PROGRAD
Pró-Reitoria de Graduação

PETIMAT
Programa de Educação Tutorial Institucional de Matemática

Notas de estudo

Métodos do Gradiente Projetado ao longo das direções viáveis e ao longo do arco de projeção

Emanuel Mendes Queiroz

Vitória da Conquista, 2024

Conteúdo

1 Introdução

1.1 O problema

Nestas notas, a discussão será realizada acerca do seguinte problema

$$\min f(x) \text{ sujeito a } x \in C, \tag{1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é uma função diferenciável e $D \subset \mathbb{R}^n$ é um conjunto de estrutura simples, isto é, convexo e fechado.

1.2 Método e estratégias

Nestas notas de aula, trataremos de duas estratégias baseadas na busca de Armijo, veja [**?**].

1.2.1 Estratégia GPA1

A primeira estratégia utilizada, que vamos chamar de GPA1, trata-se do emprego da busca de Armijo ao longo de direções viáveis do conjunto C. Neste caso, a construção da sequência $\{x_k\}$ é feita considerando-se

$$x_{k+1} = x_k + \gamma_k(z_k - x_k), \tag{2}$$

onde o parâmetro γ_k é determinado por

$$\gamma_k = 2^{-\ell(k)}, \tag{3}$$

com

$$\ell(k) = \min\left\{j \in \mathbb{Z}_+ : f(x_k - 2^{-j}(z_k - x_k)) \leq f(x_k) - \sigma 2^{-j} \nabla f(x_k)^T (x_k - z_k)\right\}. \tag{4}$$

Além disso, tomamos

$$z_k = P_C(x_k - \beta_k \nabla f(x_k)) \tag{5}$$

e a sequência $\{\beta_k\} \subset [\tilde{\beta}, \hat{\beta}]$ para algum $0 < \tilde{\beta} \leq \hat{\beta}$, onde cada β_k é calculado por meio de interpolação quadrática nesse intervalo.

1.2.2 Estratégia GPA2

A segunda estratégia será chamada de GPA2 e consiste em utilizar a busca de Armijo para determinar termos da sequência ao longo da fronteira do conjunto C. Neste caso, a determinação dos termos é dada por

$$x_{k+1} = P_C(x_k - \beta_k \nabla f(x_k)), \tag{6}$$

onde β_k é dado por

$$\beta_k = \overline{\beta} 2^{-\ell(k)}, \tag{7}$$

com

$$\ell(k) = \min\left\{j \in \mathbb{Z}_+ : f(z_{k,j}) \leq f(x_k) - \sigma \nabla f(x_k)^T (x_k - z_{k,j})\right\} \tag{8}$$

e

$$z_{k,j} = P_C(x_k - \overline{\beta} 2^{-j} \nabla f(x_k)), \tag{9}$$

para algum $\overline{\beta} > 0$ e $\sigma \in (0, 1)$.

1.2.3 Estratégia GPA3

Nesta estratégia, o comprimento de passo é determinado de maneira externa ao algoritmo e antes de ser realizada a projeção do ponto no conjunto viável. Nesse caso, β_k é dado por

$$\beta_k = \frac{\alpha_k}{\|\nabla f(x_k)\|} \tag{10}$$

onde

$$\sum_{k=0}^{\infty} \alpha_k = \infty \ \text{ e } \ \sum_{k=0}^{\infty} \alpha_k^2 < \infty.$$

2 Resultados e definições iniciais

Teorema 2.1. ***(Teorema do Valor Médio)*** *Vale o seguinte*

(a) *Se para $x, y \in \mathbb{R}^n$ uma função $F : \mathbb{R}^n \to \mathbb{R}^l$ 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 $f : \mathbb{R}^n \to \mathbb{R}$ 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*

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

Demonstração:

Lema 2.2. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no $\mathbb{R}^n$, com gradiente Lipschitz-contínuo no $\mathbb{R}^n$ com módulo $L > 0$. Então*

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

Demonstração: Pelo Teorema 2.1 (a), temos que $f(x + y) = f(x) + \int_0^1 \nabla f(x + ty)^T y\, dt$. Assim,

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

Pela Desigualdade de Cauchy-Schawrz, 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

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

□

Definição 2.3. ***(Conjunto convexo)*** *Um conjunto $C \in \mathbb{R}^n$ é chamado conjunto convexo se para quaisquer $x, y \in C$ e $\alpha \in [0,1]$, tem-se que $\alpha x + (1-\alpha)y \in C$. O ponto $\alpha x + (1-\alpha)$ onde $\alpha \in [0,1]$ chama-se combinação convexa de x e y com parâmetro α.*

Definição 2.4. ***(Função convexa)*** *Se $C \subset \mathbb{R}^n$ é um conjunto convexo, diz-se que a função $f : C \to \mathbb{R}$ é convexa em C quando para quaisquer $x, y \in C$ e $\alpha \in [0,1]$ tem-se*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y).$$

Vejamos uma interpretaçao geométrica para a desigualdade que define uma função convexa. Considere na figura abaixo a função f, os pontos $(x, f(x))$ e $(y, f(y))$ e a função φ que corresponde à equação da reta secante a esses pontos. Seja $z = \alpha x + (1-\alpha)y$ uma combinação convexa de x e y com parâmetro α.

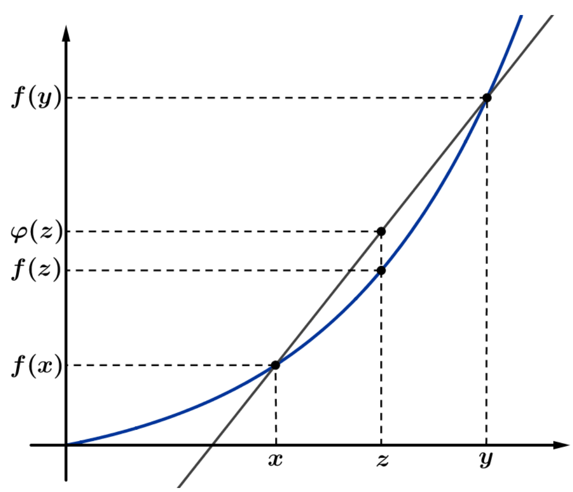

Figura 1: Exemplo de função convexa

Pela Figura 1, vemos que $f(z) < \varphi(z)$. Exibindo $\varphi(z)$ pela forma ponto-inclinação da reta, temos:

$$
\begin{aligned}
\varphi(z) &= \left(\frac{f(y)-f(x)}{y-x}\right)(z-x)+f(x) \\
&= \frac{(f(y)-f(x))(z-x)+f(x)(y-x)}{y-x} \\
&= \frac{(f(y)-f(x))(1-\alpha)(y-x)+f(x)(y-x)}{y-x} \\
&= (f(y)-f(x))(1-\alpha)+f(x) \\
&= (1-\alpha)f(y)-\alpha f(x),
\end{aligned}
$$

Donde concluímos que de fato é válida a desigualdade $f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)$ da definição.

Teorema 2.5. ***(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, $\overline{x} = P_D(x)$ se, e somente se,*

$$\overline{x} \in D, \langle x-\overline{x}, y-\overline{x}\rangle \leq 0, \ \forall y \in D. \tag{11}$$

Demonstração:

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

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

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, $L_{f,D}(c) \neq \varnothing$. Pelo Corolário **??** 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 $\overline{x}$ uma solução do problema (**??**) e $y \in D$ qualquer. Como $\overline{x} \in D$ e D é convexo, temos pela definição de combinação convexa que $(1-\alpha)\overline{x} + \alpha y = x(\alpha) \in D$ para todo $\alpha \in (0,1]$. Temos, então, que $\|x - \overline{x}\| \leq \|x - x(\alpha)\| = \|x - ((1-\alpha)\overline{x} + \alpha y)\| = \|x - \overline{x} + \alpha\overline{x} + \alpha y\|$ e $\|x - x(\alpha)\| \geq \|x - \overline{x}\| \Longrightarrow \|x - x(\alpha)\|^2 \geq \|x - \overline{x}\|^2$. Assim,

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

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

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

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

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

sendo que $y \in D$ era arbitrário.

Suponhamos agora que um certo $\overline{x}$ satisfaça (11). Então, para todo $y \in D$,

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

Agora, vamos mostrar que a projeção é única. Seja $\hat{x}$ alguma outra solução de (**??**). Usando (11) para $\overline{x}$ com $y = \hat{x} \in D$ e para $\hat{x}$ com $y = \overline{x} \in D$, temos

$$\langle x - \overline{x}, \hat{x} - \overline{x} \rangle \leq 0 \text{ e } \langle x - \hat{x}, \overline{x} - \hat{x} \rangle \leq 0 \Longrightarrow -\langle x - \hat{x}, \hat{x} - \overline{x} \rangle \leq 0.$$

Somando, obtemos que

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

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

Teorema 2.6. ***(Condições de otimalidade de primeira ordem no caso de conjunto viável convexo)*** *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e fechado, e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\overline{x} \in D$. Se $\overline{x}$ é um minimizador local de f no conjunto D, então*

$$\langle \nabla f(\overline{x}), x - \overline{x} \rangle \geq 0 \quad \forall x \in D, \tag{12}$$

ou seja, $\overline{x}$ é um ponto estacionário do problema min $f(x)$ sujeito a $x \in D$. Equivalentemente, temos

$$\overline{x} = P_D(\overline{x} - \alpha \nabla f(\overline{x})) \quad \forall \alpha \in \mathbb{R}_+. \tag{13}$$

Se f é uma função convexa, $\overline{x} \in D$ e vale (12) ou (13), então $\overline{x}$ é um minimizador global de f em D.

Demonstração:

Definição 2.7. *Um conjunto $C \subset \mathbb{R}^n$ é chamado conjunto convexo se para quaisquer $x, y \in C$ e $\alpha \in [0, 1]$ tem-se que $\alpha x + (1 - \alpha)y \in C$. O ponto $\alpha x + (1 - \alpha)y$, onde $\alpha \in [0, 1]$ é chamado de combinação convexa de x e y com parâmetro α.*

Definição 2.8. *Se $C \subset \mathbb{R}^n$ é um conjunto convexo, diz-se que a função $f : C \to \mathbb{R}$ é convexa em C quando para quaisquer $x, y \in C$ e $\alpha \in [0, 1]$ tem-se*

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Teorema 2.9. ***(Caracterizações de funções convexas diferenciáveis)*** *- Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \to \mathbb{R}$ uma função diferenciável em Ω. Se f é convexa em ω, então, para quaiquer $x, y \in \Omega$,*

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

Demonstração: Seja f convexa. Dados $x, y \in \Omega$ e $\alpha \in (0, 1]$ quaisquer, definindo $d = y - x$, temos que

$$f(x + \alpha d) = f(\alpha y + (1 - \alpha)x) \leq \alpha f(y) + (1 - \alpha)f(x) = \alpha(f(y) - f(x)) + f(x),$$

isto é,

$$\alpha(f(y) - f(x)) \geq f(x + \alpha d) - f(x).$$

Dividindo ambos os lados da desigualdade acima por $\alpha > 0$ e passando ao limite quando $\alpha \to 0^+$, obtemos

$$f(y) - f(x) \geq \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha} = \langle \nabla f(x), d \rangle = \langle \nabla f(x), y - x \rangle,$$

ou seja,

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

3 Convergência

3.1 Resultados preliminares

Definição 3.1. *Uma sequência $\{a_k\} \subset \mathbb{R}^n$ é **quasi-Fejér** convergente a um conjunto $T \subset \mathbb{R}^n$ se para todo $z \in T$ existe uma sequência $\{\varepsilon_k\} \subset \mathbb{R}_+$ tal que $\sum_{k=0}^{\infty} \varepsilon_k < \infty$ e ocorre*

$$\|a_{k+1} - z\|^2 \leq \|a_k - z\|^2 + \varepsilon_k \tag{14}$$

para todo k.

Exemplo 3.2.

Proposição 3.3. *Seja $T \subset \mathbb{R}^n$ um conjunto não-vazio e e $\{a_k\} \subset \mathbb{R}^n$ uma sequência quasi-Fejér convergente. Então:*

(i) *$\{a_k\}$ é limitada.*

(ii) *Se um ponto de acumulação $\overline{a}$ de $\{a_k\}$ pertence a T, então a sequência completa $\{a_k\}$ converge para $\overline{a}$.*

Demonstração:

(i) Fixando $z \in T$ e aplicando (14) iterativamente, temos:

$$\begin{aligned} \|a_k - z\ \|^2 &\leq \|a_{k-1} - z\|^2 + \varepsilon_{k-1} \\ &\leq \|a_{k-2} - z\|^2 + \varepsilon_{k-1} + \varepsilon_{k-2} \\ &\leq \cdots \\ &\leq \|a_0 - z\|^2 + \sum_{j=0}^{k-1} \varepsilon_j \\ &\leq \|a_0 - z\|^2 + \sum_{j=0}^{\infty} \varepsilon_j. \end{aligned}$$

Como $\sum_{j=0}^{\infty} \varepsilon_j < \infty$, podemos tomar $M = \|a_0 - z\|^2 + \sum_{j=0}^{\infty} \varepsilon_j$. Daí:

$$\|a_k - z\|^2 \leq M \Longrightarrow \|a_k - z\| \leq \sqrt{M}$$

e segue que

$$\|a_k\| = \|a_k - z + z\| \leq \|a_k - z\| + \|z\| \leq \sqrt{M} + \|z\|.$$

Logo, $\{a_k\}$ é limitada.

(ii) Sejam $\overline{a} \in T$ um ponto de acumulação de $\{a_k\}$ e $\delta > 0$ arbitrário. Seja $\{a_{\ell_k}\}$ uma subsequência de $\{a_k\}$ que converge para $\overline{a}$. Uma vez que $\{\varepsilon_k\}$ é somável, existe k_0 tal que $\sum_{j=k_0}^{\infty} \varepsilon_j < \frac{\delta}{2}$ e existe k_1 tal que $\ell_{k_1} \geq k_0$ e $\|a_{\ell_k} - \overline{a}\|^2 < \frac{\delta}{2}$ para todo $k \geq k_1$, uma vez que $\overline{a}$ é o limite de $\{a_{\ell_k}\}$. Então, para qualquer $k > \ell_{k_1}$, temos

$$\begin{aligned}
\|a_k - \overline{a}\|^2 &\leq \|a_{k-1} - \overline{a}\|^2 + \varepsilon_{k-1} \\
&\leq \|a_{k-2} - \overline{a}\|^2 + \varepsilon_{k-1} + \varepsilon_{k-2} \\
&\leq \cdots \\
&\leq \|a_{\ell_{k_1}} - \overline{a}\|^2 + \sum_{j=\ell_{k_1}}^{k-1} \varepsilon_j \\
&\leq \|a_{\ell_{k_1}} - \overline{a}\|^2 + \sum_{j=\ell_{k_1}}^{\infty} \varepsilon_j \\
&< \frac{\delta}{2} + \frac{\delta}{2} = \delta.
\end{aligned}$$

Concluímos que $\lim_{k\to\infty} a_k = \overline{a}$. $\square$

A seguir, vamos mostrar que a busca linear definida pela estratégia GPA1 sempre é bem sucedida. Iniciaremos com um fato imediato sobre direções de descida.

Proposição 3.4. *Sejam $\sigma \in (0,1)$, $x \in C$ e $v \in \mathbb{R}^n$ tais que $\langle \nabla f(x), v \rangle < 0$. Então existe $\overline{\gamma} < 1$ tal que $f(x+\gamma v) < f(x) + \sigma\gamma\langle \nabla f(x), v\rangle$ para todo $\gamma \in (0, \overline{\gamma}]$.*

Demonstração: Pela diferenciabilidade de f, para todo γ suficientemente pequeno, temos que $f(x+\gamma v) = f(x) + \gamma\langle v, \nabla f(x)\rangle + o(\gamma)$. Então,

$$\begin{aligned} f(x+\gamma v) - f(x) &= \gamma\langle \nabla f(x), v\rangle + o(\gamma) \\ &= \sigma\gamma\langle \nabla f(x), v\rangle - \sigma\gamma\langle \nabla f(x), v\rangle + \gamma\langle \nabla f(x), v\rangle + o(\gamma) \\ &= \sigma\gamma\langle \nabla f(x), v\rangle + (1-\sigma)\gamma\langle \nabla f(x), v\rangle + o(\gamma) \\ &= \sigma\gamma\langle \nabla f(x), v\rangle + \gamma[(1-\sigma)\langle \nabla f(x), v\rangle + o(\gamma)/\gamma], \end{aligned}$$

Porém, temos que para todo γ suficientemente pequeno, pela condição $\langle \nabla f(x), v\rangle < 0$, vale

$$(1-\sigma)\langle \nabla f(x), v\rangle + \frac{o(\gamma)}{\gamma} < \frac{1-\sigma}{2}\langle \nabla f(x), v\rangle < 0,$$

donde

$$f(x+\gamma v) < f(x) + \sigma\gamma\langle \nabla f(x), v\rangle.$$

Tomando $\overline{\gamma} < 1$ como o maior valor (suficientemente pequeno) que satisfaz a desigualdade acima, ela será válida também para todo $\gamma \in (0, \overline{\gamma}]$, pois se é válida para $\overline{\gamma}$, tomando $0 < \gamma < \overline{\gamma}$, o lado direito da desigualdade fica ainda maior devido à condição $\langle \nabla f(x), v\rangle < 0$. □

Lema 3.5. ***(Cota inferior para o valor do comprimento de passo dado pela regra de Armijo)*** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável $\mathbb{R}^n$, com derivada Lipschitz-contínua em $\mathbb{R}^n$ com módulo $L > 0$. Se $x_k, d_k \in \mathbb{R}^n$ satisfazem a condição $\nabla f(x_k)^T v_k < 0$, então a desigualdade $f(x_k + \gamma v_k) < f(x_k) + \sigma\gamma\nabla\langle f(x_k), v_k\rangle$ é válida para todo $\gamma \in (0, \overline{\gamma}_k]$, onde*

$$\overline{\gamma}_k = \frac{2(\sigma-1)(\nabla f(x_k)^T v_k)}{L\|v_k\|^2} > 0. \tag{15}$$

Demonstração: Pelo Lema 2.2, para todo $\alpha \in \mathbb{R}$, tem-se que

$$f(x_k+\gamma v_k) - f(x_k) - \nabla f(x_k)^T(\gamma v_k) < |f(x_k+\gamma v_k) - f(x_k) - \nabla f(x_k)^T(\gamma v_k)| \leq \frac{L}{2}\|\gamma v_k\|^2.$$

Daí, segue que

$$\begin{aligned} & f(x_k+\gamma v_k) - f(x_k) - \nabla f(x_k)^T(\gamma v_k) \leq \frac{L}{2}\gamma^2\|v_k\|^2 \\ \Longrightarrow & f(x_k+\gamma v_k) - f(x_k) \leq \nabla f(x_k)^T(\gamma v_k) + \frac{L}{2}\gamma^2\|v_k\|^2 \\ \Longrightarrow & f(x_k+\gamma v_k) - f(x_k) \leq \gamma\left(\nabla f(x_k))^T v_k + \frac{L}{2}\gamma\|v_k\|^2\right). \end{aligned}$$

Logo, para todo $\gamma \in (0, \overline{\gamma}_k]$,

$$
\begin{aligned}
& f(x_k + \gamma v_k) - f(x_k) \leq \gamma \left(\nabla f(x_k)^T v_k + \frac{L}{2} \overline{\gamma}_k \|v_k\|^2 \right) \\
\Longrightarrow & f(x_k + \gamma v_k) - f(x_k) \leq \gamma \left[\nabla f(x_k)^T v_k + \frac{L}{2} \|v_k\|^2 \cdot \left(\frac{2(\sigma - 1)(\nabla f(x_k)^T v_k)}{L\|v_k\|^2} \right) \right] \\
\Longrightarrow & f(x_k + \gamma v_k) - f(x_k) \leq \gamma (\nabla f(x_k)^T v_k + (\sigma - 1)(\nabla f(x_k)^T v_k)) \\
\Longrightarrow & f(x_k + \gamma v_k) - f(x_k) \leq \gamma (\nabla f(x_k)^T v_k + \sigma \nabla f(x_k)^T v_k - \nabla f(x_k)^T v_k) \\
\Longrightarrow & f(x_k + \gamma v_k) - f(x_k) \leq \gamma \sigma \nabla f(x_k)^T v_k,
\end{aligned}
$$

onde a segunda desigualdade segue de (15). □

No próximo resultado, vamos mostrar que a direção $z_k - x_k$ é uma direção de descida.

Proposição 3.6. *Sejam x_k e z_k definidos em (2)-(5). Então*

(i) *$x_k \in C$, para todo k.*

(ii) *Se $\nabla f(x_k) \neq 0$, então $\langle \nabla f(x_k), z_k - x_k \rangle < 0$.*

Demonstração:

(i) Aplicando indução sobre k, temos que vale para $k = 0$, pois $x_0 \in C$ pela inicialização do algoritmo. Vamos assumir que vale para um k arbitrário, ou seja, $x_k \in C$. Já que $z_k = P_C(x_k - \beta_k \nabla f(x_k))$, então $z_k \in C$. Como vale (**??**), então $\gamma_k = 2^{-\ell(k)} \in [0, 1]$. Assim, já que $x_{k+1} = x_k + \gamma_k(z_k - x_k)$, concluímos que $x_{k+1} \in C$.

(ii) Para demonstrar este item, vamos utilizar o item (b) do Teorema 2.1, segundo o qual se $P_C(u)$ é uma projeção de u em C, então vale $\langle u - P_C(u), v - P_C(u) \rangle \leq 0$, para todo $u \in \mathbb{R}^n$ e $v \in C$. Pelo item (i), temos que $x_k \in C$ para todo k. Como $z_k = P_C(x_k - \beta_k \nabla f(x_k))$, tomamos $u = x_k - \beta_k \nabla f(x_k)$, donde

$$
\begin{aligned}
& \langle x_k - \beta_k \nabla f(x_k) - P_C(x_k - \beta_k \nabla f(x_k)), x_k - P_C(x_k - \beta_k \nabla f(x_k)) \rangle \leq 0 \\
\Longrightarrow & \langle x_k - \beta_k \nabla f(x_k) - z_k, x_k - z_k \rangle \leq 0 \\
\Longrightarrow & \langle -\beta_k \nabla f(x_k) + (x_k - z_k), x_k - z_k \rangle \leq 0 \\
\Longrightarrow & -\beta_k \langle \nabla f(x_k), x_k - z_k \rangle + \langle x_k - z_k, x_k - z_k \rangle \leq 0 \\
\Longrightarrow & -\beta_k \langle \nabla f(x_k), x_k - z_k \rangle \leq 0 \\
\Longrightarrow & \langle \nabla f(x_k), x_k - z_k \rangle \geq 0 \\
\Longrightarrow & \langle \nabla f(x_k), z_k - x_k \rangle \leq 0.
\end{aligned}
$$

□

Corolário 3.7. *Se $\nabla f(x_k) \neq 0$, então γ_k está bem definido para as equações (2) - (5).*

Demonstração: Considere a *Proposição 2.4* com $x = x_k$ e $v = z_k - x_k$. Pelo item *(ii)* da *Proposição 2.5*, temos que $\langle \nabla f(x), v \rangle < 0$. Assim, a suposição da **Proposição 2.4** é válida, e existe o $\overline{\gamma}$ mencionado. Então, a desigualdade

$$\ell(k) = \min \left\{ j \in \mathbb{Z}_+ : f(x_k - 2^{-j}(z_k - x_k)) \leq f(x_k) - \sigma 2^{-j} \nabla f(x_k)^T (x_k - z_k) \right\}$$

é válida para todo j tal que $2^{-j} \leq \overline{\gamma}$, porque é válida no intervalo $(0, \overline{\gamma}]$ e $\gamma_k = 2^{-\ell(k)}$. Segue-se que tanto $\ell(k)$ quanto γ_k estão bem definidos. $\square$

Finalmente, vamos mostrar que os pontos de acumulação de $\{x_k\}$, se existirem, são estacionários.

Definição 3.8. *Um ponto $z \in C$ é estacionário para o problema (1) se, e somente se, for válido*

$$\langle \nabla f(z), x - z \rangle \geq 0 \tag{16}$$

para todo $x \in C$.

Proposição 3.9. *Seja $\{x_k\}$ a sequência definida pelas equações (2) - (5). Se $\{x_k\}$ é infinita, $\overline{x}$ é um ponto de acumulação de $\{x_k\}$ e o problema (1) tem solução, então $\overline{x}$ é estacionário para o problema (1).*

Demonstração: Uma vez que C é fechado, como $x_k \in C$ para todo k, pela Proposição 3.6(i), isso significa que $\overline{x} \in C$, pois é ponto de acumulação de C. Seja $\mathbb{N}_1 \subset \mathbb{N}$ tal que $\lim_{k\to\infty} x_k = \overline{x}$, para $k \in \mathbb{N}_1$. Observa-se que $\{\gamma_k\} \subset [0, 1]$, pois $\gamma_k = 2^{-\ell(k)}$ e pelo Corolário 3.7. Além disso, $\{\beta_k\} \subset [\tilde{\beta}, \hat{\beta}]$, onde $0 < \tilde{\beta} \leq \hat{\beta}$. Assim, podemos assumir, sem perda de generalidade, que existe $\mathbb{N}_2 \subset \mathbb{N}$ tal que se $k \in \mathbb{N}_2$ então $\lim_{k\to\infty} \gamma_k = \hat{\gamma} \in [0, 1]$ e $\lim_{k\to\infty} \beta_k = \overline{\beta} > \tilde{\beta} > 0$. Considerando as equações (2) - (4), isto é,

$$\begin{gathered} x_{k+1} = x_k + \gamma_k (z_k - x_k) \\ \gamma_k = 2^{-\ell(k)} \\ \ell(k) = \min \left\{ j \in \mathbb{Z}_+ : f(x_k - 2^{-j}(z_k - x_k)) \leq f(x_k) - \sigma 2^{-j} \nabla f(x_k)^T (x_k - z_k) \right\} \end{gathered}$$

temos que

$$\begin{aligned} & f(x_k + \gamma_k(z_k - x_k)) \leq f(x_k) - \sigma \gamma_k \nabla f(x_k)^T (x_k - z_k) \\ \Longrightarrow & f(x_{k+1}) \leq f(x_k) - \sigma \gamma_k \nabla f(x_k)^T (x_k - z_k) \\ \Longrightarrow & f(x_k) - f(x_{k+1}) \geq -\sigma \gamma_k \nabla f(x_k)^T (z_k - x_k) > 0. \end{aligned}$$

Segue que $\{f(x_k)\}$ é uma sequência decrescente. Uma vez que $\{x_k\} \subset C$ pela Proposição 3.6(i) e o problema (1) tem soluções, temos que $\{f(x_k)\}$ é limitada inferiormente, logo é convergente, donde $\lim_{k\to\infty}[f(x_k) - f(x_{k+1})] = 0$. Passando ao limite quando $k \to \infty$ e considerando $z_k = P_C(x_k - \beta_k \nabla f(x_k))$ e a continuidade de ∇f e de P_C, temos

$$\begin{aligned}
&f(x_k) - f(x_{k+1}) \geq -\sigma\gamma_k \nabla f(x_k)^T(z_k - x_k) > 0\\
\Longrightarrow &\lim_{k\to\infty}[f(x_k) - f(x_{k+1})] \geq \lim_{k\to\infty}[-\sigma\gamma_k \nabla f(x_k)^T(P_C(x_k - \beta_k\nabla f(x_k)) - x_k)] \geq 0\\
\Longrightarrow &0 \geq -\sigma\hat{\gamma}\nabla f(\overline{x})^T(P_C(\overline{x} - \overline{\beta}\nabla f(\overline{x})) - \overline{x}) \geq 0\\
\Longrightarrow &-\sigma\tilde{\gamma}\nabla f(\overline{x})^T(P_C(\overline{x} - \overline{\beta}\nabla f(\overline{x})) - \overline{x}) = 0.
\end{aligned}$$

Agora, consideremos dois casos. Primeiro, vamos supor que $\tilde{\gamma} > 0$. Seja $\overline{u} = \overline{x} - \overline{\beta}\nabla f(\overline{x})$. Então, considerando $\overline{u} = \overline{x} - \overline{\beta}\nabla f(\overline{x}))$, segue que

$$\begin{aligned}
-\sigma\hat{\gamma}\nabla f(\overline{x})^T(P_C(\overline{u}) - \overline{x}) = 0 \Longrightarrow &\nabla f(\overline{x})^T(P_C(\overline{u}) - \overline{x}) = 0\\
\Longrightarrow &\overline{\beta}^{-1}(\overline{x} - \overline{x} + \overline{\beta}\nabla f(\overline{x}))^T(P_C(\overline{u}) - \overline{x}) = 0\\
\Longrightarrow &\overline{\beta}^{-1}(\overline{x} - (\overline{x} - \overline{\beta}\nabla f(\overline{x})))^T(P_C(\overline{u}) - \overline{x}) = 0\\
\Longrightarrow &\overline{\beta}^{-1}(\overline{x} - \overline{u})^T(P_C(\overline{u}) - \overline{x}) = 0\\
\Longrightarrow &(\overline{x} - \overline{u})^T(P_C(\overline{u}) - \overline{x}) = 0\\
\Longrightarrow &\langle \overline{u} - \overline{x}, P_C(\overline{u}) - \overline{x}\rangle = 0.
\end{aligned}$$

Como $\overline{x} \in C$, pelo Teorema da Projeção, temos que:

$$\begin{aligned}
0 &\geq \langle \overline{u} - P_C(\overline{u}), \overline{x} - P_C(\overline{u})\rangle\\
&= \langle P_C(\overline{u}) - \overline{u}, P_C(\overline{u}) - \overline{x}\rangle + \langle \overline{u} - \overline{x}, P_C(\overline{u}) - \overline{x}\rangle\\
&= \langle P_C(\overline{u}) - \overline{u} + \overline{u} - \overline{x}, P_C(\overline{u}) - \overline{x}\rangle\\
&= \langle P_C(\overline{u}) - \overline{x}, P_C(\overline{u}) - \overline{x}\rangle\\
&= \|P_C(\overline{u}) - \overline{x}\|^2,
\end{aligned}$$

donde $\overline{x} = P_C(\overline{u})$. Como $\overline{\beta} > 0$, novamente pelo Teorema da Projeção, segue que

$$\begin{aligned}
\langle \overline{u} - \overline{x}, x - \overline{x}\rangle \leq 0 \Longrightarrow &\langle \overline{x} - \overline{\beta}\nabla f(\overline{x}) - \overline{x}, x - \overline{x}\rangle \leq 0\\
\Longrightarrow &-\overline{\beta}\langle \nabla f(\overline{x}), x - \overline{x}\rangle \leq 0\\
\Longrightarrow &\langle \nabla f(\overline{x}), x - \overline{x}\rangle \geq 0 \ \text{ para todo } x \in C.
\end{aligned}$$

Isto significa que $\overline{x}$ é um ponto estacionário para o problema (1).

Vamos considerar agora o caso em que $\lim_{k\to\infty}\gamma_k = \hat{\gamma} = 0$. Vamos fixar $q \in \mathbb{N}$. Uma vez que $\gamma_k = 2^{-\ell(k)}$, existe um k tal que $\ell(k) > q$. Portanto, considerando $\ell(k) = \min\left\{j \in \mathbb{Z}_+ : f(x_k - 2^{-j}(z_k - x_k)) \leq f(x_k) - \sigma 2^{-j}\nabla f(x_k)^T(x_k - z_k)\right\}$, temos

$$f(x_k + 2^{-q}(z_k - x_k)) > f(x_k) - \sigma 2^{-q}\nabla f(x_k)^T(x_k - z_k).$$

Passando ao limite quando $k \to \infty$ e definindo $\overline{z} = P_C(\overline{x} - \overline{\beta}\nabla f(\overline{x}))$, temos, para algum $q \in \mathbb{N}$ arbitrário,

$$f(\overline{x} - 2^{-q}(\overline{z} - \overline{x})) \geq f(\overline{x}) + \sigma 2^{-q}\nabla f(\overline{x})^T(\overline{z} - \overline{x})$$

Pela contrapositiva da Proposição 3.4, se $\sigma \in (0,1)$, $x \in C$ e $v \in \mathbb{R}^n$ e não existe $\overline{\gamma} < 1$ tal que $f(x+\gamma v) < f(x) + \sigma\gamma\nabla\langle f(x), v\rangle$ para todo $\gamma \in (0, \overline{\gamma}]$, então $\langle \nabla f(x), v\rangle \geq 0$. Assim, no resultado acima, obtemos que $\nabla f(\overline{x})^T(\overline{z} - \overline{x}) \geq 0$. Porém, provamos na Proposição 3.6 que $\nabla f(x_k)^T(z_k - x_k) < 0$, ou seja, passando ao limite quando $k \to \infty$, temos $\nabla f(\overline{x})^T(\overline{z} - \overline{z}) \leq 0$. Assim, deve ocorrer que $0 = \nabla f(\overline{x})^T(\overline{z} - \overline{x}) = \nabla f(\overline{x})^T(P_C(\overline{u}) - \overline{x})$. Então, a igualdade do caso anterior também é válida para este caso e concluímos que $\overline{x}$ é estacionário para o problema (1) pelos mesmos argumentos. □

Ressaltamos que nenhum dos resultados demonstrados até o momento requerem a convexidade de f. A seguir, mostraremos as propriedades de convergência no caso convexo.

3.2 Convergência no caso convexo para a estratégia GPA1

Nesta seção, vamos mostrar que quando f é convexa então a sequência gerada pela estratégia GPA1 converge para a solução do problema (1), sob a única hipótese de existência da solução.

Teorema 3.10. *Assuma que o problema (1) tenha soluções. Então ou o algoritmo que gera a sequência $\{x_k\}$, dado pelas equações (2) - (5), para em alguma iterada k, no caso em que x_k é uma solução do problema (1), ou então gera uma sequência $\{x_k\}$ infinita que converge a uma solução x^* do problema.*

Demonstração: No caso em que o algoritmo para em alguma iterada k, a condição de parada do algoritmo garante que x_k é estacionário. Uma vez que f é convexa, pontos estacionários são soluções para o problema (1). Nos próximos passos vamos assumir que o algoritmo gera uma sequência $\{x_k\}$ infinita.
Seja $\hat{x}$ qualquer solução do problema (1). Considerando $x_{k+1} = x_k + \gamma_k(z_k - x_k)$, temos que

$$\begin{aligned}
&\|x_{k+1} - x_k\|^2 + \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 \\
=&\langle x_{k+1} - x_k, x_{k+1} - x_k\rangle + \langle x_k - \hat{x}, x_k - \hat{x}\rangle - \langle x_{k+1} - \hat{x}, x_{k+1} - \hat{x}\rangle \\
=&\langle x_{k+1}, x_{k+1}\rangle - 2\langle x_{k+1}, x_k\rangle + 2\langle x_k, x_k\rangle - 2\langle x_k, \hat{x}\rangle + \langle \hat{x}, \hat{x}\rangle - \langle x_{k+1}, x_{k+1}\rangle + 2\langle x_{k+1}, \hat{x}\rangle - \langle \hat{x}, \hat{x}\rangle \\
=&2(\langle x_k, x_k\rangle - \langle x_k, \hat{x}\rangle - \langle x_{k+1}, x_k\rangle + \langle x_{k+1}, \hat{x}\rangle) \\
=&2\langle x_k - x_{k+1}, x_k - \hat{x}\rangle \\
=&2\langle x_k - (x_k + \gamma_k(z_k - x_k)), x_k - \hat{x}\rangle \\
=&-2\gamma_k\langle z_k - x_k, x_k - \hat{x}\rangle \\
=&2\gamma_k\langle z_k - x_k, \hat{x} - x_k\rangle.
\end{aligned}$$

Utilizando novamente o teorema da projeção, isto é, $\langle P_C(u) - u, v - P_C(u)\rangle \geq 0$ para todo

$u \in \mathbb{R}^n$ e todo $v \in C$ e considerando $z_k = P_C(x_k - \beta_k \nabla f(x_k))$, temos que como $\hat{x} \in C$,

$$\begin{aligned}
0 \leq & \langle z_k - (x_k - \beta_k \nabla f(x_k)), \hat{x} - z_k \rangle \\
= & \langle z_k - x_k + \beta_k \nabla f(x_k), (\hat{x} - x_k) + (x_k - z_k) \rangle \\
= & \langle z_k - x_k + \beta_k \nabla f(x_k), \hat{x} - x_k \rangle + \langle z_k - x_k + \beta_k \nabla f(x_k), x_k - z_k \rangle \\
= & \langle z_k - x_k, \hat{x} - x_k \rangle - \beta_k \langle \nabla f(x_k), x_k - \hat{x} \rangle + \langle z_k - x_k + \beta_k \nabla f(x_k), x_k - z_k \rangle,
\end{aligned}$$

donde segue que

$$\begin{aligned}
\langle z_k - x_k, \hat{x} - x_k \rangle \geq & \beta_k \langle \nabla f(x_k), x_k - \hat{x} \rangle - \langle z_k - x_k + \beta_k \nabla f(x_k), x_k - z_k \rangle \\
\geq & \beta_k [f(x_k) - f(\hat{x})] + \langle z_k - x_k + \beta_k \nabla f(x_k), z_k - x_k \rangle \\
\geq & \langle z_k - x_k + \beta_k \nabla f(x_k), z_k - x_k \rangle \\
= & \langle z_k - x_k, z_k - x_k \rangle + \beta_k \langle \nabla f(x_k), z_k - x_k \rangle \\
= & \|z_k - x_k\|^2 + \beta_k \langle \nabla f(x_k), z_k - x_k \rangle,
\end{aligned}$$

onde utilizamos o Teorema (2.9) e o fato de que $x_k \in C$ pela Proposição (3.6)(i) na segunda desigualdade e a otimalidade de $\hat{x}$ e a positividade de β_k na terceira desigualdade. Além disso, de (2) obtemos

$$x_{k+1} = x_k + \gamma_k (z_k - x_k) \Longrightarrow z_k - x_k = \frac{x_{k+1} - x_k}{\gamma_k} \Longrightarrow \|z_k - x_k\|^2 = \left\| \frac{x_{k+1} - x_k}{\gamma_k} \right\|^2 = \gamma_k^{-2} \|x_{k+1} - x_k\|^2.$$

Combinando os três resultados obtidos anteriormente, isto é,

$$\|x_{k+1} - x_k\|^2 + \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 = 2\gamma_k \langle z_k - x_k, \hat{x} - x_k \rangle$$

$$\langle z_k - x_k, \hat{x} - x_k \rangle \geq \|z_k - x_k\|^2 + \beta_k \langle \nabla f(x_k), z_k - x_k \rangle$$
$$\|z_k - x_k\|^2 = \gamma_k^{-2} \|x_{k+1} - x_k\|^2$$

segue que

$$\begin{aligned}
\|x_{k+1} - x_k\|^2 + \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 & \geq 2\gamma_k \left[\|z_k - x_k\|^2 + \beta_k \langle \nabla f(x_k), z_k - x_k \rangle \right] \\
& = 2\gamma_k \left[\gamma_k^{-2} \|x_{k+1} - x_k\|^2 + \beta_k \langle \nabla f(x_k), z_k - x_k \rangle \right] \\
& = 2\gamma_k^{-1} \|x_{k+1} - x_k\|^2 + 2\gamma_k \beta_k \langle \nabla f(x_k), z_k - x_k \rangle.
\end{aligned}$$

Rearranjando a desigualdade acima, obtemos

$$\begin{aligned}
& \|x_{k+1} - x_k\|^2 + \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 \geq 2\gamma_k^{-1} \|x_{k+1} - x_k\|^2 + 2\gamma_k \beta_k \langle \nabla f(x_k), z_k - x_k \rangle \\
\Longrightarrow & \|x_{k+1} - \hat{x}\|^2 \leq \|x_{k+1} - x_k\|^2 + \|x_k - \hat{x}\|^2 - 2\gamma_k^{-1} \|x_{k+1} - x_k\|^2 - 2\gamma_k \beta_k \langle \nabla f(x_k), z_k - x_k \rangle \\
\Longrightarrow & \|x_{k+1} - \hat{x}\|^2 \leq \|x_k - \hat{x}\|^2 + (1 - 2\gamma_k^{-1}) \|x_{k+1} - x_k\|^2 - 2\gamma_k \beta_k \langle \nabla f(x_k), z_k - x_k \rangle \\
\Longrightarrow & \|x_{k+1} - \hat{x}\|^2 \leq \|x_k - \hat{x}\|^2 - 2\gamma_k \beta_k \langle \nabla f(x_k), z_k - x_k \rangle, \qquad (10)
\end{aligned}$$

onde utilizamos o fato de que $\gamma_k \in [0, 1]$ na última desigualdade.
Agora, vamos considerar especificamente a forma que os parâmetros de comprimento de passo γ_k's são determinados. Por (3) e (4), temos que, para todo p,

$$f(x_p + \gamma_p(z_p - x_p)) \leq f(x_p) - \sigma\gamma_p \nabla f(x_p)^T (x_p - z_p)$$
$$\implies -\sigma\gamma_p \nabla f(x_p)^T (z_p - x_p) \leq f(x_p) - f(x_p + \gamma_p(z_p - x_p))$$
$$\implies -\sigma\gamma_p \nabla f(x_p)^T (z_p - x_p) \leq f(x_p) - f(x_{p+1})$$

Multiplicando a desigualdade anterior por $2\beta_p/\sigma$, obtemos

$$-2\beta_p\gamma_p \nabla f(x_p)^T (z_p - x_p) \leq \frac{2\beta_p}{\sigma}[f(x_p) - f(x_{p+1})].$$

Seja $\varepsilon_p = -2\beta_p\gamma_p \nabla f(x_p)^T (z_p - x_p) > 0$, uma vez que $z_p - x_p$ é direção de descida e β_p e γ_p são positivos. Considerando que $\{f(x_p)\}$ é decrescente pela Proposição (3.9) e que $\{\beta_p\} \subset [\tilde{\beta}, \hat{\beta}]$, obtemos

$$\varepsilon_p \leq \frac{2\beta_p}{\sigma}[f(x_p) - f(x_{p+1})] \leq \frac{2\hat{\beta}}{\sigma}[f(x_p) - f(x_{p+1})].$$

Somando a desigualdade acima com p entre 0 e k e levando em conta que $\hat{x}$ é solução do problema (1),

$$\sum_{p=0}^{k} \varepsilon_p \leq \frac{2\hat{\beta}}{\sigma}[f(x_0) - f(x_{k+1})] \leq \frac{2\hat{\beta}}{\sigma}[f(x_0) - f(\hat{x})].$$

Passando ao limite quando $k \to \infty$, obtemos que $\sum_{p=0}^{\infty} \varepsilon_p < \infty$. Levando em conta (10) e que $\varepsilon_p = -2\beta_p\gamma_p \nabla f(x_p)^T (z_p - x_p)$, chegamos em

$$\|x_{k+1} - \hat{x}\|^2 \leq \|x_k - \hat{x}\|^2 + \varepsilon_k. \tag{18}$$

Seja S^* o conjunto de soluções do problema (1). Uma vez que $\hat{x}$ é um elemento arbitrário de S^* e $\sum_{p=0}^{\infty} \varepsilon_p < \infty$, (18) implica que $\{x_k\}$ é quasi-Fejér convergente para S^*. Já que S^* é não vazio por hipótese, segue da Proposição (3.3)(i) que $\{x_k\}$ é limitada e, portanto, possui pontos de acumulação. Pela Proposição (3.9), todos os pontos de acumulação de $\{x_k\}$ são estacionários. Pela convexidade de f, esses pontos são soluções para o problema (1), ou seja, pertencem a S^*. Pela Proposição (3.3)(ii), a sequência completa $\{x_k\}$ converge para uma solução do problema (1). □

3.3 Convergência para a estratégia GPA2

Nesta seção, será feita a análise correspondente à estratégia GPA2, que seguem diretamente dos resultados para a estratégia GPA1, desenvolvidos nas duas seções anteriores. Sem assumir a convexidade de f, os resultados a seguir são válidos.

Proposição 3.11. *Para todo $x \in C$ e $z \in \mathbb{R}^n$, a função $\varphi : (0, +\infty) \to \mathbb{R}$ definida por*

$$\varphi(\beta) = \frac{\|P_C(x + \beta z) - x\|}{\beta}$$

é monótona não crescente.

Demonstração: Vamos tomar dois escalares β_1 e β_2 com $\beta_1 > 0$ e $\beta_2 > \beta_1$ e mostrar que

$$\frac{\|P_C(x+\beta_2 z)-x\|}{\beta_2} \leq \frac{\|P_C(x+\beta_1 z)-x\|}{\beta_1}. \tag{19}$$

Para isso, vamos fazer algumas mudanças de variáveis estratégicas. Considere $y = \beta_1 z$, $\gamma = \beta_2/\beta_1$, $a = x + y$ e $b = x + \gamma y$. Assim, (19) pode ser reescrita como

$$\begin{aligned}
&\frac{\|P_C(x+\beta_2 z)-x\|}{\beta_2} \leq \frac{\|P_C(x+\beta_1 z)-x\|}{\beta_1} \\
\Longrightarrow &\left\| P_C\left(x + \frac{\beta_2}{\beta_1}\beta_1 z\right) - x \right\| \leq \frac{\beta_2}{\beta_1}\|P_C(x+\beta_1 z)-x\| \\
\Longrightarrow &\|P_C(b)-x\| \leq \gamma\|P_C(a)-x\|. \qquad (17)
\end{aligned}$$

Vamos analisar separadamente os possíveis casos. Se $P_C(a) = x$, temos $\|P_C(b)-x\| = 0 \Longrightarrow P_C(b) = x$ e a desigualdade (17) é válida. Se $a \in C$, então $P_C(a) = a = x + y$, de modo que (17) se torna $\|P_C(b)-x\| \leq \gamma\|y\| = \|b-x\|$, que também é válido para uma projeção ortogonal. Finalmente, se $P_C(a) = P_C(b)$, então (17) também é válida, pois $\gamma > 1$. Portanto, é suficiente mostrar que (17) é válida no caso em que $P_C(a) \neq P_C(b)$, $P_C(a) \neq x$, $P_C(b) \neq x$ e $a \notin C$.

Sejam H_a e H_b os dois hiperplanos que são ortogonais a $P_C(b) - P_C(a)$ e passam por $P_C(a)$ e $P_C(b)$, respectivamente. Uma vez que, pelo Teorema da Projeção $\langle P_C(b)-P_C(a), b-P_C(b)\rangle \geq 0$ e $\langle P_C(b) - P_C(a), a - P_C(a)\rangle \leq 0$, temos que nem a e nem b se encontram estritamente entre os dois hiperplanos H_a e H_b. Além disso, x se encontra do mesmo lado de H_a que a está, então $x \notin H_a$. Sejam s_a e s_b as interseções da reta $\{x + \alpha(b-x); \alpha \in \mathbb{R}\}$ com H_a H_b, respectivamente. Seja w a interseção da reta $\{x + \alpha(P_C(a)-x); \alpha \in \mathbb{R}\}$ com H_b. A Figura (2) representa o caso que está sendo analisado.

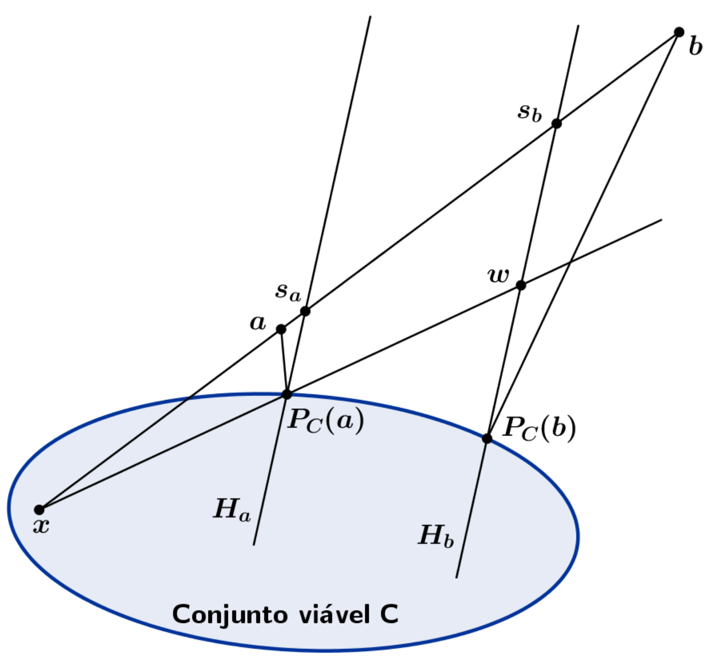

Figura 2: Interpretação geométrica para a Proposição (3.11)

Note que os triângulos $\triangle s_a x P_C(a)$ e $\triangle s_b x w$ são semelhantes, de modo que

$$\frac{\|s_b - x\|}{\|s_a - x\|} = \frac{\|w - x\|}{\|P_C(a) - x\|}.$$

Além disso, os segmentos $\overline{wP_C(b)}$ e $\overline{P_C(b)P_C(a)}$ são ortogonais, ou seja, o triângulo $\triangle P_C(a)P_C(b)w$ é retângulo, donde

$$\|w - P_C(a)\| \geq \|P_C(b) - P_C(a)\|.$$

Assim, considerando os fatos acima e desde que $y \neq 0$, temos

$$\begin{aligned}
\gamma = \gamma\frac{\|y\|}{\|y\|} = \frac{\|\gamma y\|}{\|y\|} = \frac{\|b - x\|}{\|a - x\|} \geq \frac{\|s_b - x\|}{\|s_a - x\|} &= \frac{\|w - x\|}{\|P_C(a) - x\|} \\
&= \frac{\|w - P_C(a)\| + \|P_C(a) - x\|}{\|P_C(a) - x\|} \\
&\geq \frac{\|P_C(b) - P_C(a)\| + \|P_C(a) - x\|}{\|P_C(a) - x\|} \\
&\geq \frac{\|P_C(b) - x\|}{\|P_C(a) - x\|},
\end{aligned}$$

donde obtemos (17), o que conclui a prova. □

Proposição 3.12. *Seja $\{x_k\}$ a sequência gerada por (6) - (9) Então:*

- ***(i)*** $\{x_k\} \subset C$.
- ***(ii)*** *Se $\nabla f(x_k) \neq 0$, então $\langle \nabla f(x_k), x_k(\beta_k) - x_k \rangle < 0$, onde $x_k(\beta_k) = x_{k+1} = P_C(x_k - \beta_k \nabla f(x_k))$.*

Demonstração:

(i) Como $x_0 \in C$ e $x_{k+1} = P_C(x_k - \beta_k \nabla f(x_k))$, é imediato que $\{x_k\} \subset C$.

(ii) Como $x_k \in C$ do item anterior, pelo Teorema da Projeção, temos que

$$\begin{aligned}
&\langle x_k - x_k(\beta_k), (x_k - \beta_k \nabla f(x_k)) - x_k(\beta_k) \rangle \leq 0 \\
\Longrightarrow &\langle x_k - x_k(\beta_k), x_k - x_k(\beta_k) \rangle \leq \beta_k \langle \nabla f(x_k), x_k - x_k(\beta_k) \rangle \\
\Longrightarrow &0 < \frac{\|x_k - x_k(\beta_k)\|^2}{\beta_k} \leq \langle \nabla f(x_k), x_k - x_k(\beta_k) \rangle \qquad (18) \\
\Longrightarrow &\langle \nabla f(x_k), x_k(\beta_k) - x_k \rangle < 0,
\end{aligned}$$

onde assumimos que x_k não é estacionário, ou seja, $\|x_k - x_k(\beta_k)\| \neq 0$.

Proposição 3.13. *β_k está bem definido por (6) - (9), isto é, dado $\sigma \in (0, 1)$, para todo $x \in C$, existe $\beta^* > 0$ tal que*

$$f(x(\beta)) \leq f(x) - \sigma \nabla f(x)^T (x - x(\beta)), \qquad (19)$$

para todo $\beta \in (0, \beta^*]$.

Demonstração: Se x é estacionário, então $\|x - x(\beta)\| = 0 \Longrightarrow x = x(\beta)$, teremos $f(x) \leq f(x) - \sigma \nabla f(x)^T (x - x) \Longrightarrow 0 \leq 0$ e a conclusão vale com β^* sendo qualquer escalar positivo. Vamos assumir então que x não é estacionário, ou seja, $\|x - x(\beta)\| \neq 0$ para todo $\beta > 0$. Pelo Teorema do Valor Médio, temos que

$$\begin{aligned}
f(x) - f(x(\beta)) =& \langle \nabla f(\xi_\beta), x - x(\beta) \rangle \\
=& \langle \nabla f(\xi_\beta) - \nabla f(x) + \nabla f(x), x - x(\beta) \rangle \\
=& \langle \nabla f(x), x - x(\beta) \rangle + \langle \nabla f(\xi_\beta) - \nabla f(x), x - x(\beta) \rangle,
\end{aligned}$$

onde ξ_β pertence ao segmento de extremidades x e $x(\beta)$. Portanto, podemos reescrever a equação (19) da seguinte forma:

$$\begin{aligned}
& \langle \nabla f(x), x - x(\beta) \rangle + \langle \nabla f(\xi) - \nabla f(x), x - x(\beta) \rangle \geq \sigma \langle \nabla f(x), x - x(\beta) \rangle \\
\Longrightarrow & (1 - \sigma) \langle \nabla f(x), x - x(\beta) \rangle \geq - \langle \nabla f(\xi) - \nabla f(x), x - x(\beta) \rangle \\
\Longrightarrow & (1 - \sigma) \langle \nabla f(x), x - x(\beta) \rangle \geq \langle \nabla f(x) - \nabla f(\xi), x - x(\beta) \rangle \qquad (20)
\end{aligned}$$

Da equação (18) e da Proposição (3.11), temos que, para todo $\beta \in (0, 1]$, vale

$$\langle \nabla f(x), x - x(\beta) \rangle \geq \|x - x(\beta)\| \frac{\|x - x(\beta)\|}{\beta} \geq \|x - x(\beta)\| \|x - x(1)\|.$$

Portanto, a equação (20) será satisfeita para todo $\beta \in (0, 1]$ que satisfizer

$$(1 - \sigma) \|x - x(1)\| \geq \left\langle \nabla f(x) - \nabla f(\xi_\beta), \frac{x - x(\beta) \rangle}{\|x - x(\beta)\|} \right\rangle.$$

Para um β suficientemente pequeno, temos que $x(\beta) \to x$ e, portanto, $\xi_\beta \to x$, donde o lado direito da desigualdade acima tende a zero, de modo que existe algum β^* que a satisfaz. Portanto, as equações (20) e (19) serão satisfeitas para todo $\beta \in (0, \beta^*]$.

Proposição 3.14. *Se o problema (1) possui soluções e $\overline{x}$ é um ponto de acumulação de $\{x_k\}$, então $\overline{x}$ é um ponto estacionário para o problema.*

Demonstração: A proposição anterior garante que β_k está bem definido como um número positivo para todo k. Seja $\overline{x}$ um ponto de acumulação de $\{x_k\}$ e seja $\mathbb{N}_1 \subset \mathbb{N}$ com $\mathbb{N}_1$ infinito tal que $\lim_{k\to\infty} x_k = \overline{x}$ para $k \in \mathbb{N}_1$. Considerando as equações (6) - (9) e a Proposição (3.12)(ii), temos que

$$\begin{aligned}
& f(x_k(\beta_k)) \leq f(x_k) - \sigma \nabla f(x_k)^T (x_k - x_k(\beta_k)) \\
\Longrightarrow & f(x_k) - f(x_k(\beta_k)) \geq \sigma \nabla f(x_k)^T (x_k - x_k(\beta_k)) > 0 \\
\Longrightarrow & f(x_k) - f(x_{k+1}) > 0,
\end{aligned}$$

donde $\{f(x_k)\}$ é monótona decrescente e, portanto, $f(x_k) \to f(\overline{x})$ para $k \in \mathbb{N}_1$. Sem perda de generalidade, podemos supor que $\lim_{k\to\infty} \beta_k = \hat{\beta} \geq 0$, para $k \in \mathbb{N}_2 \subset \mathbb{N}$ com $\mathbb{N}_2$ infinito.

Vamos considerar dois casos possíveis.

Se ocorrer que $\lim_{k\to\infty}\beta_k = \hat{\beta}$ para algum $\hat{\beta} > 0$, então da equação (18) e da Proposição (3.11) segue que, para todo k suficientemente grande

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq \sigma\nabla f(x_k)^T(x_k - x_{k+1}) \\ &\geq \frac{\sigma\|x_k - x_{k+1}\|^2}{\beta_k} \\ &= \frac{\sigma\beta_k\|x_k - x_k(\beta_k)\|^2}{\beta_k^2} \\ &\geq \frac{\sigma\hat{\beta}\|x_k - x_k(\overline{\beta})\|^2}{\overline{\beta}^2}. \end{aligned}$$

Passando ao limite quando $k \to \infty$, temos

$$\frac{\sigma\hat{\beta}\|\overline{x} - \overline{x}(\overline{\beta})\|}{\overline{\beta}^2} \leq 0 \Longrightarrow \|\overline{x} - \overline{x}(\overline{\beta})\| = 0 \Longrightarrow \overline{x} = \overline{x}(\overline{\beta}).$$

Portanto, $\overline{x}$ é estacionário.

Vejamos agora o caso em que $\lim_{k\to\infty}\beta_k = 0$. Fixando $q \in \mathbb{N}$, como $\beta_k = \overline{\beta}2^{-\ell(k)}$, existe algum $k \in \mathbb{N}_2$ suficientemente grande tal que $\ell(k) > q$, de modo que o teste de Armijo falha para $\ell(k) = q$, ou seja, $j \geq 1$ em (8) e

$$f(x_k) - f(z_{k,q}) < \sigma\nabla f(x_k)^T(x_k - z_{k,q}). \tag{23}$$

Além disso, para todo $k \in \mathbb{N}_2$, x_k não pode ser estacionário. De fato, se fosse x_k estacionário, teríamos $x_k = P_C(x_k - \overline{\beta}\nabla f(x_k)) = z_{k,0}$ pelo Teorema ?? e a desigualdade de Armijo seria satisfeita para $j = 0$, pois

$$f(x_k) - f(z_{k,0}) \geq \sigma\nabla f(x_k)^T(x_k - z_{k,0}) \Longrightarrow f(x_k) \geq f(z_{k,0}), \tag{24}$$

o que implicaria que $\beta_k = \overline{\beta}$, absurdo. Portanto, deve ocorrer

$$\|x_k - z_{k,q}\| > 0. \tag{25}$$

Pelo Teorema do Valor Médio, temos

$$\begin{aligned} f(x_k) - f(z_{k,q}) &= \nabla f(\xi_k)^T(x_k - z_{k,q}) \\ &= (\nabla f(\xi_k) - \nabla f(x_k) + \nabla f(x_k))^T(x_k - z_{k,q}) \\ &= \nabla f(x_k)^T(x_k - z_{k,q}) + (\nabla f(\xi_k) - \nabla f(x_k))^T(x_k - z_{k,q}), \end{aligned}$$

onde ξ_k pertence ao segmento de extremidades x_k e $z_{k,q}$. Combinando a equação acima com (23), obtemos

$$\begin{aligned} &\nabla f(x_k)^T(x_k - z_{k,q}) + (\nabla f(\xi_k) - \nabla f(x_k))^T(x_k - z_{k,q}) < \sigma\nabla f(x_k)^T(x_k - z_{k,q}) \\ \Longrightarrow &(1-\sigma)\nabla f(x_k)^T(x_k - z_{k,q}) < (\nabla f(x_k) - \nabla f(\xi_k))^T(x_k - z_{k,q}). \end{aligned} \tag{26}$$

Utilizando (18) e a Proposição (3.11), obtemos

$$\nabla f(x_k)^T(x_k - z_{k,q}) \geq \frac{\|x_k - z_{k,q}\|^2}{\overline{\beta}2^{-q}} = \frac{\|x_k - z_{k,q}\|}{\overline{\beta}2^{-q}}\|x_k - z_{k,q}\| \geq \frac{\|x_k - z_{k,0}\|}{\overline{\beta}}\|x_k - z_{k,q}\|. \quad (27)$$

Combinando (25), (26) e (27) e utilizando a desigualdade de Cauchy-Schwarz, obtemos

$$(1-\sigma)\left(\frac{\|x_k - z_{k,0}\|}{\overline{\beta}}\|x_k - z_{k,q}\|\right) < (\nabla f(x_k) - \nabla f(\xi_k))^T(x_k - z_{k,q})$$

$$\Longrightarrow \frac{1-\sigma}{\overline{\beta}}\|x_k - z_{k,0}\|\|x_k - z_{k,q}\| < \|\nabla f(x_k) - \nabla f(\xi_k)\|\|x_k - z_{k,q}\|$$

$$\Longrightarrow \frac{1-\sigma}{\overline{\beta}}\|x_k - z_{k,0}\| < \|\nabla f(x_k) - \nabla f(\xi_k)\|. \quad (28)$$

Uma vez que $\beta_k \to 0$, $z_{k,q} \to x_k$ e $x_k \to \overline{x}$ quando $k \to \infty$, segue que $\xi_k \to \overline{x}$ quando $k \to \infty$. Portanto, passando ao limite quando $k \to \infty$ em (28), obtemos

$$\|x_k - \overline{x}(\overline{\beta})\| \leq 0 \Longrightarrow \overline{x} = \overline{x}(\overline{\beta}),$$

donde segue que $\overline{x}$ é estacionário. □

Teorema 3.15. *Assuma que o problema (1) tenha solução. Então ou o algoritmo que gera a sequência $\{x_k\}$, dado pelas equações (6) - (9), para em alguma iterada k, no caso em que x_k é uma solução do problema (1), ou então gera uma sequência $\{x_k\}$ infinita que converge a uma solução x^* do problema.*

Demonstração: O resultado para o caso em que a sequência é finita segue do critério de parada e da convexidade de f. Para o caso de uma sequência infinita, primeiramente observamos que os cálculos feitos na demonstração do Teorema (3.10) até a desigualdade (10) não utilizam a maneira específica pela qual β_k's e γ_k's são definidos, então eles são válidos para a sequência que estamos considerando aqui, onde agora $\gamma_k = 1$ para todo k e β_k é dado por (7) - (9). Portanto, para toda solução $\hat{x}$ do problema (1), nós temos

$$\|x_{k+1} - \hat{x}\|^2 \leq \|x_k - \hat{x}\|^2 + \varepsilon_k, \quad (29)$$

onde

$$\varepsilon_k = 2\beta_k \nabla f(x_k)^T[x_k - P_C(x_k - \beta_k \nabla f(x_k))]. \quad (30)$$

Note que $\varepsilon_k \geq 0$ para todo k pela Proposição (3.12)(ii) e pela positividade de β_k. A seguir, vamos mostrar que $\sum_{k=0}^{\infty} \varepsilon_k < \infty$. Considerando as equações (6) - (9), particularmente o critério de busca no arco, e a definição de ε_k, temos

$$f(x_{k+1}) \leq f(x_k) - \sigma \nabla f(x_k)^T(x_k - x_{k+1})$$

$$\Longrightarrow \sigma \nabla f(x_k)^T[x_k - P_C(x_k - \beta_k \nabla f(x_k))] \leq f(x_k) - f(x_{k+1})$$

$$\Longrightarrow 2\beta_k \nabla f(x_k)^T[x_k - P_C(x_k - \beta_k \nabla f(x_k))] \leq \left(\frac{2\beta_k}{\sigma}\right)(f(x_k) - f(x_{k+1}))$$

$$\Longrightarrow \varepsilon_k \leq \left(\frac{2\beta_k}{\sigma}\right)(f(x_k) - f(x_{k+1})) \leq \left(\frac{2\overline{\beta}}{\sigma}\right)(f(x_k) - f(x_{k+1})) \quad (31)$$

Somando (31) com k entre 0 e j, e considerando a otimalidade de $\hat{x}$, obtemos

$$\sum_{k=0}^{j} \varepsilon_k \leq \left(\frac{2\overline{\beta}}{\sigma}\right)(f(x_0) - f(x_{k+1})) \leq \left(\frac{2\overline{\beta}}{\sigma}\right)(f(x_0) - f(\hat{x})).$$

Passando ao limite quando $k \to \infty$, obtemos

$$\sum_{k=0}^{\infty} \varepsilon_k \leq \left(\frac{2\overline{\beta}}{\sigma}\right)(f(x_0) - f(\hat{x})) < \infty.$$

Considerando (29), segue que, assim como no Teorema (3.10), a sequência $\{x_k\}$ é quasi-Fejér convergente para para o conjunto de soluções, e então a Proposição (3.3)(i) implica que $\{x_k\}$ é limitada, então possui pontos de acumulação. Pela Proposição (3.14) e pela convexidade de f, todos esses pontos de acumulação são estacionários e, portanto, soluções para o problema (1). Finalmente, a Proposição (3.3)(ii) implica que toda a sequência $\{x_k\}$ converge para uma solução. □