

Avanços nos Métodos de Descida

Coordenada por Blocos

Vitaliano S. Amaral – UFPI

vitalianoamaral@ufpi.edu.br

vitalianoamaral.github.io

XIII Jornada de Matemática da UFPI - JMatUFPI



Setembro de 2025

Métodos de Descida Coordenada por Blocos BCD

Considere o seguinte problema:

$$\text{Minimizar } f(x) \text{ sujeito a } x \in \Omega \subset \mathbb{R}^n. \quad (1)$$

Basicamente um método BCD para resolver o Problema (1) consiste em:

Método BCD

Dados $k \leftarrow 0$, $x^0 \in \Omega$.

Repetir os seguintes passos:

Passo 1. Dado x^k escolher $I_k \subset \{1, 2, \dots, n\}$;

Passo 2. Atualizar apenas as coordenadas do bloco escolhido:

$$x_i^{k+1} = \begin{cases} x_i^k, & i \notin I_k, \\ x_i^{k+1} \text{ (valor que gera decréscimo em } f\text{,)} & i \in I_k. \end{cases}$$

Definir $k \leftarrow k + 1$ e voltar ao Passo 1.

Métodos de Descida Coordenada por Blocos BCD

- ▶ Os métodos BCD estão entre os primeiros de decomposição de variáveis, onde um dos primeiros algoritmos foi descrito para a minimização de funções quadráticas por Gauss e Seidel.
- ▶ Ideia central: decompor um grande problema em subproblemas menores.
- ▶ Por um bom tempo, esses métodos receberam pouca atenção por parte dos pesquisadores. Alguns motivos:
 - ▶ Baixo desempenho prático.
 - ▶ Poucos desafios teóricos.
- ▶ A situação mudou drasticamente nas últimas décadas. Dentre vários motivos, destacamos:
 - ▶ Aplicações em problemas de grande porte.
 - ▶ Desenvolvimento de novas teorias.

Um dos pioneiros na retomada do estudo dos métodos BCD foi o trabalho de Beck e Tetrushvili¹.

Problema considerado:

$$\min f(x) \quad \text{s.a.} \quad x \in \mathbb{R}^n.$$

- ▶ Os autores assumiram f convexa,
- ▶ Gradiente parcial $\nabla_{(i)} f$ Lipschitz.

Método do Gradiente Coordenado por Blocos - BCGD

Tome $x^0 \in \mathbb{R}^n$ e L_i ($i = 1, \dots, p$) constantes de Lipschitz do gradiente parcial $\nabla_{(i)} f$.

Passo 1. Tome $x^{k,0} = x^k$ e defina recursivamente:

$$x^{k,i} = x^{k,i-1} - \frac{1}{L_i} u_i \nabla_{(i)} f(x^{k,i-1}), \quad i = 1, \dots, p$$

Passo 2. Defina $x^{k+1} = x^{k,p}$

¹A. Beck, and L. Tetrushvili (2013). - *On the convergence of block coordinate descent type methods*. SIAM Journal on Optimization, 23(4):20372060.

Convergência do método BCGD

Assumindo f convexa e condições de Lipschitz nos gradientes parciais, os autores provaram que a sequência $\{f(x^k)\}$ satisfaz as seguintes condições:

► Para todo $k = 0, 1, 2, \dots$

$$f(x^k) - f(x^{k+1}) \geq \frac{1}{4L_{\max}(1 + pL^2/L_{\min}^2)} \|\nabla f(x^k)\|^2$$

$$f(x^k) - f(x^*) \leq 4L^2(1 + pL^2/L_{\min}^2)R^2(x^0) \frac{1}{k + 8/p}, \quad k = 0, 1, 2, \dots,$$

onde L_{\max} e L_{\min} são as constantes de Lipschitz máxima e mínima dos p -blocos, respectivamente,

$$R(x_0) \equiv \max_{x \in \mathbb{R}^n} \max_{x^* \in X^*} \{\|x - x^*\| : f(x) \leq f(x_0)\}$$

e $f(x^*)$ é um valor ótimo de f .

Método BCGD no \mathbb{R}^2

Tome $x^0 \in \mathbb{R}^2$ e L_i ($i = 1, 2$) constantes de Lipschitz de $\nabla_{(i)} f$.

Passo 1. Tome $x^{k,0} = x^k = (x_1^{k,0}, x_2^{k,0})$ e defina recursivamente:

$$x^{k,1} = x^{k,0} - \frac{1}{L_1} e_1 \frac{\partial f(x^{k,0})}{\partial x_1} = \left(x_1^{k,0} - \frac{1}{L_1} \frac{\partial f(x^{k,0})}{\partial x_1}, x_2^{k,0} \right)$$

$$x^{k,2} = x^{k,1} - \frac{1}{L_2} e_2 \frac{\partial f(x^{k,1})}{\partial x_2} = \left(x_1^{k,1}, x_2^{k,1} - \frac{1}{L_2} \frac{\partial f(x^{k,1})}{\partial x_2} \right)$$

Passo 2. Defina

$$x^{k+1} = x^{k,2} = \left(x_1^{k,0} - \frac{1}{L_1} \frac{\partial f(x^{k,0})}{\partial x_1}, x_2^{k,1} - \frac{1}{L_2} \frac{\partial f(x^{k,1})}{\partial x_2} \right).$$

- Vale destacar que os subproblemas resolvidos no Passo 1 são resolvidos na reta real.

Teste do BCGD em Problemas de Mínimos Quadrados

Considere a função objetivo $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, tal que:

$$f(x) = \frac{1}{2} \|Ax - b\|^2, \quad \text{onde } A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}_{2 \times 2} \text{ e } b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{2 \times 1}$$

Temos $\nabla f(x) = A^T(Ax - b)$ e $L_i = 9$. Consideramos $x^0 = (1, \frac{2}{3})$.

It.	Bloco i = 1		Bloco i = 2		
	$U_1 \nabla_1 f(x^{k,0})$	$x^{k,1}$	$U_2 \nabla_2 f(x^{k,1})$	$x^{k,2} = x^{k+1}$	$f(x^{k+1})$
1	(4.6666, 0)	(0.4814, 0.6666)	(0, 2.2592)	(0.4814, 0.4156)	0.120578
2	(1.0699, 0)	(0.3625, 0.4156)	(0, 0.5285)	(0.3625, 0.3569)	0.006290
3	(0.2406, 0)	(0.3358, 0.3569)	(0, 0.1279)	(0.3625, 0.3569)	0.000329
4	(0.0500, 0)	(0.3303, 0.3426)	(0, 0.0346)	(0.3303, 0.3388)	0.000032
5	(0.0068, 0)	(0.3295, 0.3388)	(0, 0.0123)	(0.3295, 0.3374)	0.000016
6	(-0.0024, 0)	(0.3298, 0.3374)	(0, 0.0065)	(0.3298, 0.3367)	0.000012
7	(-0.0040, 0)	(0.3302, 0.3367)	(0, 0.0047)	(0.3302, 0.3362)	0.000009





Tabela: Resultados do Método BCGD em f para $k = 0, 1, 2, 3, 4, 5, 6, 7$.

Um outro trabalho importante na retomada dos estudos método BCD foi o trabalho de S. J. Wright².

- ▶ O artigo descreve os fundamentos da abordagem, variantes, extensões e propriedades de convergência, sobretudo no caso de funções convexas.
- ▶ Dá atenção especial a estruturas que surgem frequentemente em aprendizado de máquina, mostrando implementações eficientes de versões aceleradas.
- ▶ Além disso, apresenta variantes paralelas e discute suas propriedades de convergência em diferentes modelos de execução.

²S. J. Wright. - *Coordinate descent algorithms*. Mathematical Programming, 151(1):3-34, 2015

Recentemente surgiram vários trabalhos sobre os métodos CD.
Por exemplo, os trabalhos:

-  V. S. Amaral; R. Andreani; E. J. G. Birgin; D. S. Marcondes and J. M. Martínez. - *On complexity and convergence of high-order coordinate descent algorithms for smooth nonconvex box-constrained minimization*. Journal of Global Optimization (2022): 1-35
-  E. G. Birgin, and J. M. Martínez. - *Block coordinate descent for smooth nonconvex constrained minimization*. Computational Optimization and Applications 83.1 (2022): 1-27.
-  I. Necoara and F. Chorobura. - *Random Coordinate Descent Methods for Nonseparable Composite Optimization*. SIAM Journal on Optimization. 33, 2160-2190 (2023)
-  Amaral, V. (2025). - *A partially derivative-free cyclic block coordinate descent method for nonseparable composite optimization*. Mathematical Modelling and Analysis, 30(3), 535-552.

O restante dessa apresentação é dedicada aos seguintes trabalhos:



V. S. Amaral; R. Andreani; E. J. G. Birgin; D. S. Marcondes and J. M. Martínez. - *On complexity and convergence of high-order coordinate descent algorithms for smooth nonconvex box-constrained minimization*. Journal of Global Optimization (2022): 1-35



Amaral, V. (2025). - *A partially derivative-free cyclic block coordinate descent method for nonseparable composite optimization*. Mathematical Modelling and Analysis, 30(3), 535-552.

No primeiro trabalho³ foi considerado o seguinte problema:

$$\text{Minimize } f(\mathbf{x}) \quad \text{s.a. } \mathbf{x} \in \Omega, \quad (2)$$

onde $\Omega \subset \mathbb{R}^n$ é dado por $\Omega = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\}$, e $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$ são tais que $\mathbf{l} < \mathbf{u}$.

- ▶ f com derivadas primeiras contínuas em Ω .
- ▶ $\mathbf{g}_P(\mathbf{x}) = P_\Omega(\mathbf{x} - \nabla f(\mathbf{x})) - \mathbf{x}$ para todo $\mathbf{x} \in \Omega$, onde P_Ω é a projeção euclidiana em Ω .
- ▶ $\mathbf{g}_{P,I}(\mathbf{x}) \in \mathbb{R}^n$ definida por

$$[\mathbf{g}_{P,I}(\mathbf{x})]_i = \begin{cases} [\mathbf{g}_P(\mathbf{x})]_i, & \text{se } i \in I, \\ 0, & \text{se } i \notin I. \end{cases}$$

- ▶ $M_{\bar{\mathbf{x}}}(\cdot)$ uma aproximação de f em torno de $\bar{\mathbf{x}}$.

³V. S. Amaral; R. Andreani; E. J. G. Birgin; D. S. Marcondes and J. M. Martínez.
- *On complexity and convergence of high-order coordinate descent algorithms for smooth nonconvex box-constrained minimization*. Journal of Global Optimization (2022): 1-35

Introduzimos o seguinte método para resolver o Problema (10):

Algoritmo 1.

Suponha $p \in \{1, 2, 3, \dots\}$, $\alpha > 0$, $\sigma_{\min} > 0$, $\tau_2 \geq \tau_1 > 1$, $\theta > 0$ e $x_0 \in \Omega$ sejam fornecidos. Inicialize $k \leftarrow 0$ e $\sigma_0 \leftarrow 0$.

Passo 1. Escolha um conjunto não vazio $I_k \subseteq \{1, \dots, n\}$.

Passo 2. Calcule $x^{\text{trial}} \in \Omega$, $x_i^{\text{trial}} = x_i^k$ para todos $i \notin I_k$ tal que

$$M_{x^k}(x^{\text{trial}}) + \sigma_k \|x^{\text{trial}} - x^k\|^{p+1} \leq M_{x^k}(x^k)$$

e

$$\|P_\Omega [x^{\text{trial}} - \nabla (M_{x^k}(x) + \sigma_k \|x - x^k\|^{p+1}) \big|_{x=x^{\text{trial}}}] - x^{\text{trial}}\| \leq \theta \|x^{\text{trial}} - x^k\|^p.$$

Passo 3. Se

$$f(x^{\text{trial}}) \leq f(x^k) - \alpha \|x^{\text{trial}} - x^k\|^{p+1},$$

definir $x^{k+1} = x^{\text{trial}}$, $\sigma_{k+1} = \sigma_k$ e voltar ao Passo 1. Caso contrário, atualize $\sigma_k \leftarrow \max\{\sigma_{\min}, \tau \sigma_k\}$ com $\tau \in [\tau_1, \tau_2]$ e voltar ao Passo 2.

Boa definição

Assumimos que:

Para todo $x \in \mathbb{R}^n$ existe $L > 0$ tal que

$$\|\nabla f(x) - \nabla M_{\bar{x}}(x)\| \leq L\|x - \bar{x}\|^p,$$

$$M_{\bar{x}}(\bar{x}) = f(\bar{x}) \quad \text{e} \quad f(x) \leq M_{\bar{x}}(x) + L\|x - \bar{x}\|^{p+1}.$$

Para satisfazer as condições anteriores, $M_{\bar{x}}(\cdot)$ pode ser Taylor de ordem p de f ao redor de \bar{x} se as derivadas de ordem p de f satisfazem condições de Lipschitz ou a própria f .

Provamos que para $\sigma_k \geq L + \alpha$, o ponto x^{trial} satisfaz

$$f(x^{\text{trial}}) \leq f(x^k) - \alpha\|x^{\text{trial}} - x^k\|^{p+1} \quad (3)$$

e

$$\|\nabla g_{p,I_k}(x^{\text{trial}})\| \leq (L + \tau_2(L + \alpha)(p + 1) + \theta)\|x^{\text{trial}} - x^k\|^p, \quad (4)$$

Garantindo a boa definição do método.

Convergência

Provamos que a sequência $\{x^k\}$ gerada pelo Algoritmo 1 satisfaz as seguintes condições:

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0, \quad (5)$$

$$\lim_{k \rightarrow \infty} \|\nabla g_{P, I_k}(x^{k+1})\| = 0, \quad (6)$$

e

$$\lim_{k \rightarrow \infty} \|\nabla g_{P, I_k}(x^k)\| = 0. \quad (7)$$

Observe que os resultados anteriores não garantem

$$\lim_{k \rightarrow \infty} \|g_P(x^k)\| = 0. \quad (8)$$

Para provar (8) foi necessário uma suposição adicional em relação a escolha dos blocos.

Suposição 1.

Existe $\bar{m} < +\infty$ tal que, para todo $i \in \{1, \dots, n\}$:

1. Existe $k \leq \bar{m}$ tal que $i \in I_k$;
2. Para qualquer $k \in \mathbb{N}$, se $i \in I_k$, então existe $m \leq \bar{m}$ tal que $i \in I_{k+m}$.

Note que a Suposição 1 nos permite escolher o bloco de coordenadas em cada iteração de diversas formas, basta que, a cada \bar{m} iterações, todos os blocos sejam escolhidos pelo menos uma vez.

Com essa suposição adicional, mostramos que a sequência $\{x^k\}$ satisfaz,

$$\lim_{k \rightarrow \infty} \|\nabla g_P(x^k)\| = 0. \quad (9)$$

Além disso, se $x^* \in \Omega$ for um ponto de acumulação de $\{x^k\}$, então temos que $\|\nabla g_P(x^*)\| = 0$.

A complexidade do método

Dado as estimativas $f_{\text{target}} < f(x^0)$ e $\epsilon > 0$ dados. Provamos que o número de iterações k para obter

$$f(x^{k+1}) \leq f_{\text{target}} \text{ ou } |[g_P(x^{k+1})]_i| \leq \epsilon \text{ para todo } i \in I_k$$

é no máximo

$$\frac{f(x^0) - f_{\text{target}}}{c \epsilon^{\frac{p+1}{p}}},$$

onde c depende apenas de α , τ_2 , L , p , e θ .

Podemos observar que a estimativa anterior para o gradiente projetado se refere apenas às coordenadas escolhidas no bloco.

No entanto, obter $|[g_P(x^k)]_i| \leq \epsilon$ para todos $i \notin I_k$ é mais difícil, pois para este propósito, é necessário que as iterações consecutivas estejam suficientemente próximas.

Para isso, foi assumido a seguinte suposição.

A complexidade do método

Existe $L_g > 0$ tal que para todo $i = 1, \dots, n$ e $x, z \in \Omega$,

$$|[g_P(x)]_i - [g_P(z)]_i| \leq L_g \|x - z\|.$$

Agora, considerando a suposição adicional anterior, provamos que o número máximo de iterações necessário para obter $f(x^k) \leq f_{\text{target}}$ ou $|[g_P(x^k)]_i| \leq \varepsilon$ para todo $i = 1, \dots, n$ é

$$\frac{f(x^0) - f_{\text{target}}}{c(\frac{1}{2})^{\frac{p+1}{p}}} \varepsilon^{-\frac{p+1}{p}} + \frac{f(x^0) - f_{\text{target}}}{\alpha(\frac{1}{2\bar{m}L_g})^{p+1}} \varepsilon^{-(p+1)} + 1.$$

Assim, a complexidade obtida é da ordem $O(\varepsilon^{-(p+1)})$.

A ordem da complexidade obtida chama a atenção, pois piora quando aumentamos a ordem de regularização, ao contrário do que ocorre em métodos tradicionais.

Alguns pontos sobre

$$\frac{f(x^0) - f_{\text{target}}}{c(\frac{1}{2})^{\frac{p+1}{p}}} \varepsilon^{-\frac{p+1}{p}} + \frac{f(x^0) - f_{\text{target}}}{\alpha(\frac{1}{2\bar{m}L_g})^{p+1}} \varepsilon^{-(p+1)} + 1.$$

1. p cresce \implies a complexidade piora, o contrário dos métodos tradicionais.
2. \bar{m} cresce \implies o segundo termo da expressão acima cresce.
3. \bar{m} crescer junto com n se o tamanho dos subproblemas permanecer limitado.
4. O tipo de escolha do bloco pode influenciar na complexidade.

No trabalho⁴ foi considerado o seguinte problema:

$$\text{Minimizar } F(x) := f(x) + h(x) \text{ sujeito a } x \in \mathbb{R}^n, \quad (10)$$

onde f diferenciável e $h : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa.

Quando h é a função indicadora de um conjunto convexo C , então o Problema (10) é equivalente a resolver o problema de minimizar $F(x) = f(x)$ restrito ao conjunto convexo C .

Muitos problemas de otimização em aplicações do mundo real são frequentemente de grande escala e alta dimensão, o que os torna desafiadores de resolver, especialmente quando há derivadas difíceis de calcular.

Neste trabalho foi proposto uma versão dos métodos BCD parcialmente livre de derivadas.

⁴V. S. Amaral. - *A partially derivative-free cyclic block coordinate descent method for nonseparable composite optimization*. Mathematical Modelling And Analysis, v. 30, p. 535-552, 2025.

Para o desenvolvimento do método proposto, foi considerado a função

$$\varphi_f : \mathbb{R}^n \times [0, 1] \rightarrow \mathbb{R}^n \text{ onde } \lim_{\lambda \rightarrow 0} \varphi_f(x, \lambda) = \nabla f(x). \quad (11)$$

A função em (11) pode ser definida como $\varphi_f = \nabla f$. No entanto, essa definição não é recomendada quando o cálculo do gradiente de f é computacionalmente proibitivo. Nesses casos, é preferível definir φ_f como uma aproximação do gradiente ∇f , que pode ser obtida com um custo computacional menor.

A função φ_f pode ser definido por uma das seguintes maneiras:

- $\varphi_f(x, \lambda) = \left[\frac{f(x + \lambda e_1) - f(x)}{\lambda}, \dots, \frac{f(x + \lambda e_n) - f(x)}{\lambda} \right].$
- $\varphi_f(x, \lambda) = \left[\frac{f(x) - f(x - \lambda e_1)}{\lambda}, \dots, \frac{f(x) - f(x - \lambda e_n)}{\lambda} \right].$
- $\varphi_f(x, \lambda) = \left[\frac{f(x + \lambda e_1) - f(x - \lambda e_1)}{2\lambda}, \dots, \frac{f(x + \lambda e_n) - f(x - \lambda e_n)}{2\lambda} \right].$

Seja $x^0 \in \mathbb{R}^n$; para cada $i = 1, \dots, q$, uma matriz simétrica semidefinida positiva $B_{(i)}(x^0) \in \mathbb{R}^{n_i \times n_i}$; tome $\alpha \in (0, 1)$, $\epsilon \in (0, 1)$, $\sigma_0 \geq 1$ e $F_{\text{target}} \in \mathbb{R}$. Inicialize $k \leftarrow 0$.

Etapa 1: Escolha $\lambda_k \in \left[0, \frac{\epsilon}{\sigma_k \sqrt{n}}\right]$, e considere $\varphi_f(x^k, \lambda_k)$.

Etapa 2: Ponha $x^{k,0} = x^k$ e, para cada $i = 1, \dots, q$, calcule

$x^{k,i} = x^{k,i-1} + U_i s_{(i)}^k$, onde $s_{(i)}^k \in \mathbb{R}^{n_i}$ é solução de

$$\min_{s \in \mathbb{R}^{n_i}} \langle U_i^T \varphi_f(x^{k,i-1}, \lambda_k), s \rangle + \frac{1}{2} \langle B_{(i)}(x^{k,i-1}) s, s \rangle + h(x^{k,i-1} + U_i s) - h(x^{k,i-1}) + \frac{\sigma_k}{2} \|s\|^2.$$

Etapa 3: Se

$$\left\| \sum_{i=1}^q U_i s_{(i)}^k \right\|_{\infty} < \frac{\epsilon}{\sigma_k} \quad \text{ou} \quad F(x^{k,q}) \leq F_{\text{target}},$$

pare e declare $x^{k,q}$ como uma solução aceitável. Caso contrário, vá para a Etapa 4.

Etapa 4: Se

$$F(x^{k,q}) \leq F(x^k) - \frac{\alpha}{\sigma_k} \epsilon^2,$$

faça $k \leftarrow k + 1$, defina $x^{k+1} = x^{k,q}$ e $\sigma_{k+1} = \sigma_k$, escolha $B_{(i)}(x^{k+1}) \in \mathbb{R}^{n_i \times n_i}$ simétrica s.d.p. e volte à Etapa 1. Caso contrário, defina $\sigma_k \leftarrow 2\sigma_k$ e volte à Etapa 1.

Boa definição

Para garantir a boa definição do método foi considerado que:

Suposição 2.

Existem $L, M \in (0, +\infty)$ e $\beta \in (0, 1]$ tais que para todo $x, y \in \mathbb{R}^n$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 + \frac{M}{\beta + 1} \|y - x\|^{\beta+1} \quad (12)$$

e

$$\|\nabla f(x) - \varphi_f(x, \lambda)\| \leq \frac{\sqrt{n}L}{2} \lambda + \frac{\sqrt{n}M}{\beta + 1} \lambda^\beta \quad (13)$$

Se $f := g_1 + g_2$, onde ∇g_1 L -Lipschitz e ∇g_2 M -Hölder com expoente β então (12) e (13) são válidas, pois

$$g_1(y) \leq g_1(x) + \langle \nabla g_1(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 \quad \forall x, y \in \mathbb{R}^n$$

e

$$g_2(y) \leq g_2(x) + \langle \nabla g_2(x), y - x \rangle + \frac{M}{\beta + 1} \|y - x\|^{\beta+1} \quad \forall x, y \in \mathbb{R}^n.$$

Boa definição

Além das suposições anterior foi considerado que: existe $\bar{B} \geq 1$ tal que $\|B_{(i)}(x^{k,i-1})\| \leq \bar{B}$ para todos os $i \in \{1, \dots, q\}$ e k .

Assim, provamos que se

$$\|s^k\|_\infty \geq \frac{\epsilon}{\sigma_k} \text{ e } \sigma_k \geq \left[\frac{2L + q\bar{B}}{(1-\alpha)} + \frac{2q(Mn^{\frac{1-\beta}{2}} + M)}{(\beta+1)(1-\alpha)} \epsilon^{\beta-1} \right]^{\frac{1}{\beta}}$$

então

$$F(x^{k,q}) \leq F(x^k) - \alpha \frac{\sigma_k}{2} \|s^k\|_\infty^2 \quad (14)$$

e

$$F(x^{k,q}) \leq F(x^k) - \frac{\alpha}{\sigma_k} \epsilon^2. \quad (15)$$

Garantindo assim a boa definição do método.

Complexidade

- ▶ Números de iterações é limitado superiormente por:

$$\frac{c(F(x^0) - F_{\text{target}})}{\alpha} e^{-\frac{\beta+1}{\beta}}.$$

- ▶ Número de avaliações de f e seus subdiferenciais é limitado superiormente por:

$$\frac{c(F(x^0) - F_{\text{target}})}{\alpha} e^{-\frac{\beta+1}{\beta}} + \log_2 \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right).$$

com

$$\sigma_{\max} = 2\epsilon^{\frac{\beta-1}{\beta}} \max \left\{ \sigma_{\min}, \left[\frac{2L + q\bar{B}}{(1-\alpha)} + \frac{2q(Mn^{\frac{1-\beta}{2}} + M)}{(\beta+1)(1-\alpha)} \right]^{\frac{1}{\beta}} \right\}.$$

O método BCDC-Dfree pode resolver problemas:

- ▶ **Caso diferenciável:** se $h \equiv 0$

$$\min_x f(x).$$

- ▶ **o problema do lasso**

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1 \right\}, \lambda > 0, A \in \mathbb{R}^{s \times n}, \quad (16)$$

considerando $h(x) = \lambda \|x\|_1$, $f(x) = \frac{1}{2} \|Ax - b\|^2$.

- ▶ **o problema**

$$\min \{ \|Ax - b\|^2 : 0 \leq x_i \leq 1, i = 1, 2, \dots, n \} \quad (17)$$

onde $A \in \mathbb{R}^{s \times n}$ e $b \in \mathbb{R}^s$.

Basta minimizar $F(x) = f(x)$ com $f(x) = \delta_{\Omega}(x)$,

$\Omega = \{x \in \mathbb{R}^n : 0 \leq x_i \leq 1, i = 1, 2, \dots, n\}$, $f(x) = \|Ax - b\|^2$.

- **Problema de mínimos quadrados penalizado por norma L_p**
($1 < p < 2$):

$$\min F(x) = \frac{1}{2} \|A(x) - b\|_2^2 + \frac{\lambda}{p} \|\Phi(x)\|_p^p,$$

onde $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é diferenciável (não necessariamente linear),
 $b \in \mathbb{R}^m$, e Φ é operador linear.

Detalhes em:



A. Bernigaud, S. Gratton and E. Simon, A non-linear conjugate gradient in dual space for L_p -norm regularized non-linear least squares with application in data assimilation, *Numerical Algorithms*, **95**, 471–497 (2024).

<https://doi.org/10.1007/s11075-023-01578-x>

Aplicamos nosso método proposto para resolver alguns problemas em que ambos foi considerados como um caso particular de

$$F(x) = \frac{1}{2} \|A(x) - b\|_2^2 + \frac{\lambda}{p} \|\Phi(x)\|_p^p, \quad (18)$$

com $1 < p < 2$, A diferenciável e Φ linear.

Nos exemplos a seguir, consideramos $B_{(i)} \equiv 0$, $n = 10$, $\alpha = 0,5$, $\epsilon = 5 \times 10^{-5}$, $F_{\text{target}} = 0,5$ e a função $\varphi_f : \mathbb{R}^{10} \times \mathbb{R} \rightarrow \mathbb{R}^{10}$ definida por

$$\varphi_f(x, \lambda) = \left[\frac{f(x + \lambda e_1) - f(x)}{\lambda}, \dots, \frac{f(x + \lambda e_{10}) - f(x)}{\lambda} \right].$$

Exemplo Numérico: Influência de σ_0

A desigualdade em (15) mostra que valores muito grandes de σ_0 podem gerar pequenas reduções na função objetivo, demandando mais iterações para alcançar uma solução aceitável.

Consideramos o seguinte problema de mínimos quadrados lineares, penalizado por uma norma L_p , com $1 < p < 2$:

$$\min F(x) = \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda}{p} \|\Phi(x)\|_p^p,$$

onde $A \in \mathbb{R}^{10 \times 10}$ e $b \in \mathbb{R}^{10}$ são fixados.

No algoritmo utilizamos $f(x) = 0.5 \|Ax - b\|^2 + 5 \times 10^{-5} \sum_{i=1}^{10} |x_i|^{3/2}$, $h(x) = 0$ e $x^0 = 0 \in \mathbb{R}^{10}$.

λ_k	N. Blocos	σ_0	N. Iterações	Approx. F^*
$\frac{\epsilon}{\sigma_k \sqrt{10}}$	10	1	10	0.4731
	10	500	40	0.3021
	10	1000	83	0.3003

Tabela: Comportamento do BCDC-Dfree para diferentes σ_0 .

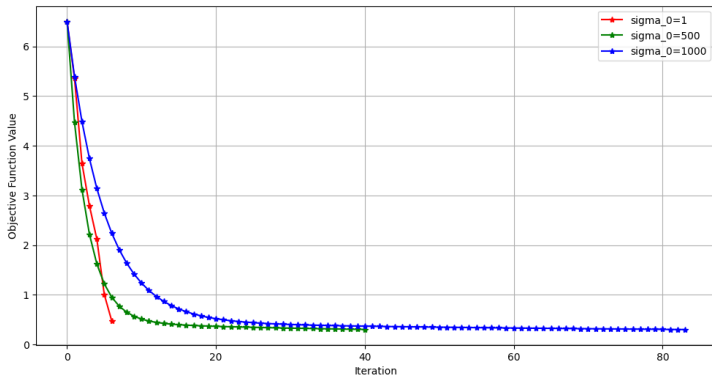


Figura: Evolução do BCDC-Dfree para diferentes valores de σ_0 .

Exemplos Numéricos com A e b Aleatórios

Consideramos o Problema (10) com f dado em (18), onde $A \in \mathbb{R}^{10 \times 10}$ e $b \in \mathbb{R}^{10}$ são gerados aleatoriamente, Φ é a função identidade e h é nulo. Foram avaliadas 100 instâncias distintas, todas com x^0 aleatório, comparando a abordagem com 10 blocos e com um único bloco.

Blocos	σ_0	Iterações Médias	Média-F
10	1	11.55	0.4329
1	1	17.56	0.4682

Tabela: BCDC-Dfree com A e b aleatórios.

Em seguida, σ_0 foi sorteado no intervalo $[1, 1000]$, mantendo os demais dados inalterados.

Blocos	σ_0	Iterações Médias	Média-F
10	rand()	805.45	0.4988
1	rand()	809.02	0.4988

Tabela: BCDC-Dfree com A , b e σ_0 aleatórios.

Exemplo com A Não Linear

Agora, consideramos o caso em que A em (18) é não linear, definido por

$$A(x) = (A_1(x), \dots, A_n(x)), \quad A_i(x) = \frac{x_i}{1 + x_{i+1}^2}, \quad i = 1, \dots, n-1,$$

com $x_{n+1} = x_1$. Tomamos $\Phi(x) = Bx$, com $B \in \mathbb{R}^{m \times n}$ aleatório e b também aleatório. Foram avaliadas 100 instâncias, todas com x^0 aleatório e σ_0 sorteado em $[1, 100]$.

Blocos	σ_0	Iterações Médias	Média-F
10	rand()	648.96	0.5008
1	rand()	664.61	0.5008

Tabela: BCDC-Dfree para o caso com A não linear.