

Decomposição aditiva de matrizes

Considere novamente problemas de estimação cujo parâmetro tem a estrutura de matriz. Modelamos anteriormente o caso em que o parâmetro tem posto pequeno. Existem aplicações em que vale a pena modelar o parâmetro da forma

$$B^* = \Gamma^* + S^*,$$

onde Γ^* é uma matriz de posto r e S^* é uma matriz com no máximo s entradas não-nulas. são matrizes $\mathbb{R}^{d_1 \times d_2}$. Neste projeto iremos implementar métodos iterativos para achar a solução do seguinte estimador de mínimos quadrados regularizado:

$$\begin{aligned} & \underset{[\Gamma, S] \in (\mathbb{R}^{d_1 \times d_2})^2}{\text{minimize}} && f(\Gamma, S) + \lambda \|\Gamma\|_N + \tau \|S\|_1, \\ & \text{sujeito à} && \|\Gamma\|_\infty \leq a, \end{aligned}$$

onde

$$f(\Gamma, S) := \frac{1}{2n} \sum_{i=1}^n \{y_i - \text{tr}(X_i^\top (\Gamma + S))\}^2,$$

$(y_1, X_1), \dots, (y_n, X_n)$ é uma amostra de labels/features em $\mathbb{R} \times \mathbb{R}^{d_1 \times d_2}$. Acima, $\lambda, \tau, a > 0$ são hyper-parâmetros,

$$\|\Gamma\|_N := \sum_{j=1}^{\min\{d_1, d_2\}} \sigma(B),$$

é a *norma nuclear* da matriz Γ , onde $\sigma_1(\Gamma) \geq \dots \geq \sigma_{\min\{d_1, d_2\}}(\Gamma)$ são os valores singulares de B . Além disso,

$$\|\Gamma\|_\infty := \max_{(j,k) \in [d_1] \times [d_2]} |\Gamma_{jk}|,$$

é a *norma* ℓ_∞ da matriz Γ e

$$\|S\|_1 := \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} |S_{jk}|,$$

é a *norma* ℓ_1 da matriz S .

Exemplos de aplicação de variações deste problema incluem [multi-task learning](#) e [análise multiômica](#).

Soft-Thresholding matricial

Os métodos que veremos para os problemas anteriores se baseam na idéia de *splitting*. Por exemplo, note que o o problema de otimização em PCA robusta exige otimizar em duas variáveis $[\Gamma, S]$. Entretanto otimizar simultaneamente é mais difícil do que otimizar alternadamente.

Precisaremos computar os operadores proximais das normas nuclear e ℓ_1 . O *operador proximal* da norma $\lambda \|\cdot\|_N$ com restrição $\|\Gamma\|_\infty \leq a$ é

$$\begin{aligned} P_N(W, \lambda, a) = & \underset{\Gamma \in \mathbb{R}^{d_1 \times d_2}}{\text{armin}} && \frac{1}{2} \|W - \Gamma\|_F^2 + \lambda \|\Gamma\|_N, \\ & \text{sujeito à} && \|\Gamma\|_\infty \leq a. \end{aligned}$$

$P_N(W, \lambda, a)$ tem fórmula explícita. Dados $\gamma \in \mathbb{R}$, defina

$$S_a(\gamma, \lambda) := \text{sign}(\gamma) \cdot \min \{\max\{\gamma - \lambda, 0\}, a\}.$$

Acima, $\text{sign}(\gamma)$ é o sinal de γ . Seja agora a decomposição de valores singulares (SVD) de W :

$$W = U \cdot D(\gamma_1, \dots, \gamma_r) \cdot V^\top.$$

Então,

$$P_N(W, \lambda, a) = U \cdot D(S_a(\gamma_1, \lambda), \dots, S_a(\gamma_r, \lambda)) \cdot V^\top,$$

chamado de matriz *constraint soft-thresholding* de W .

O *operador proximal* da norma $\tau\|\cdot\|_1$ é

$$P_1(W, \tau) \in \operatorname{argmin}_{S \in \mathbb{R}^{d_1 \times d_2}} \left\{ \frac{1}{2} \|W - S\|_F^2 + \tau \|S\|_1 \right\}.$$

$P_1(W, \tau)$ também tem fórmula explícita. Dados $\gamma \in \mathbb{R}$, defina

$$S(\gamma, \lambda) := \operatorname{sign}(\gamma) \cdot \max\{\gamma - \lambda, 0\}.$$

$P_1(W, \tau)$ é a matriz $d_1 \times d_2$ cuja $[j, k]$ -ézima entrada é $S(W_{j,k}, \tau)$.

```
In [1]: import matplotlib
import numpy as np
import scipy
from scipy.linalg import svd
from scipy.stats import ortho_group
import seaborn as sns
import matplotlib.pyplot as plt
import numpy.linalg as la
import math
```

Exercício 1:

Construa as funções `soft(x,l)` que retorna $S(x, \lambda)$ e `csoft(x,l,a)` que retorna $S_a(x, \lambda)$.

```
In [2]: #Escreva o código aqui

def soft(x, l):
    return np.sign(x) * np.maximum(np.abs(x) - l, 0)

def csoft(x, l, a):
    return np.sign(x) * np.minimum(np.maximum(np.abs(x) - l, 0), a)
```

Exercício 2: Gerando dados

Construa uma função `data_genG(n,d1,d2,r,Gamma_mag)` que toma d_1, d_2 , posto r e número positivo `Gamma_mag` e retorna a matriz Γ^* de dimensão $d_1 \times d_2$ e posto r com $\|\Gamma\|_\infty$ no máximo Γ_{mag}/\sqrt{n} . Ao gerar esta matriz Γ^* , use a função `scipy.stats.ortho_group` para construir duas matrizes U e V aleatórias ortogonais de dimensões $d_1 \times r$ e $d_2 \times r$ respectivamente; compute $U \cdot D(\Gamma_{mag}, \dots, \Gamma_{mag}) \cdot V^\top$ e finalmente, normalize esta matriz entrada por entrada para satisfazer $\|\Gamma^*\| \leq \Gamma_{mag}/\sqrt{n}$.

```
In [3]: #Escreva o código aqui
def data_genG(n:int, d1:int, d2:int, r:int, Gamma_mag):
    U = ortho_group.rvs(d1)[:r, :r]
    V = ortho_group.rvs(d2)[:r, :r]

    D = np.diag([Gamma_mag] * r)

    G = U @ D @ V.T

    max_abs_value = np.max(np.abs(G))
    scaling_factor = Gamma_mag / (np.sqrt(n) * max_abs_value)
    G *= scaling_factor

    return G
```

```
In [4]: #Exemplo:
n=100
d_1=10
d_2=10
r=5
Gamma_mag=10
Gamma_true = data_genG(n,d_1,d_2,r,Gamma_mag)
Gamma_true.shape, np.max(abs(Gamma_true)), Gamma_true,
```

```
Out[4]: ((10, 10),
1.0,
array([[ 0.29775719,  0.72806023,  0.12885487, -0.2545146 ,  0.75187686,
        -0.01603354, -0.67102822,  0.4050779 ,  0.2618715 , -0.48949832],
       [-0.11996512, -0.47374424, -0.06059274, -0.82101441, -0.16253627,
        0.8098959 , -0.15696674,  0.71762541, -0.11423484, -0.05251154],
       [-0.2679059 ,  0.00111047,  0.24338491,  0.44396351, -0.78195254,
        -0.31371313, -0.28829448,  0.62957207,  0.15159404, -0.3354399 ],
       [-0.25566648, -0.30939157, -0.09122687, -0.08674152, -0.00960433,
        0.10043277, -0.22975644,  0.0698954 , -0.03202864,  0.14589224],
       [ 0.30213009,  0.54167209,  0.36690311, -0.45403835,  0.06449996,
        -0.25479071,  0.50630757, -0.44867795, -0.11130366,  0.04687795],
       [-0.13289269,  0.53953048,  0.65668047, -0.17944117, -0.60785618,
        -0.56962641, -0.18818681,  0.39551727,  0.09339723, -0.33068837],
       [-0.51320403, -0.39072905, -0.04877195,  0.46696399, -0.31492642,
        -0.40749553, -0.22362699, -0.23064219, -0.01231467,  0.30447594],
       [-0.04046158,  0.14759389, -0.15713829,  0.4773914 ,  0.47953978,
        -0.34835611, -0.31441261, -0.36140432,  0.12540311,  0.07353889],
       [ 0.58215257,  0.02840588, -0.49807498,  0.98947909,  0.0653644 ,
        0.43379277,  0.21528811,  0.46217953,  0.23550711, -0.55100178],
       [ 0.50678268,  0.18054449,  0.12435125,  0.06233785, -0.55307723,
        0.1583128 ,  1.          , -0.03285364, -0.11234328, -0.18120356]]))
```

Exercício 3: Gerando dados

Construa uma função `data_genS(d1,d2,s,S_mag)` que toma d_1, d_2 , esparsidade $s \in [d_1d_2]$ e número positivo `S_mag` e retorna a matriz $d_1 \times d_2$ com s entradas iguais a S_{mag} e outras entradas iguais a zero. Escolha as entradas não-nulas aleatoriamente.

```
In [5]: #Escreva o código aqui
def data_genS(d1, d2, s, S_mag):
    S_2 = np.zeros((d1, d2))

    indices = np.random.choice(d1 * d2, size=s, replace=False)

    np.put(S_2, indices, S_mag)

    return S_2
```

```
In [6]: #Exemplo:
d_1=12
d_2=5
s=10
S_mag=10
S_true = data_genS(d_1,d_2,s,S_mag)
S_true
```

```
Out[6]: array([[10.,  0.,  0.,  0.,  0.],
       [ 0.,  0.,  0.,  0.,  0.],
       [ 0.,  0., 10.,  0.,  0.],
       [ 0.,  0.,  0.,  0.,  0.],
       [10.,  0.,  0., 10.,  0.],
       [ 0.,  0.,  0.,  0.,  0.],
       [ 0.,  0., 10.,  0., 10.],
       [ 0.,  0.,  0.,  0.,  0.],
       [ 0.,  0., 10.,  0., 10.],
       [ 0.,  0.,  0., 10.,  0.],
       [ 0.,  0.,  0.,  0.,  0.],
       [ 0.,  0., 10.,  0.,  0.]])
```

Exercício 4:

Construa uma função `data_genXe(n,d_1,d_2,Gamma_true,S_true,sd)` que toma $n, d_1, d_2, \Gamma^*, S^*$ e um número positivo `sd` e constrói os dados `X`, uma lista de n matrizes X_i de dimensão $d_1 \times d_2$ independentes cujas entradas são iid normais padrão. A função também retorna o vetor `y` de dimensão n cujas coordenadas satisfazem

$$y_i = tr(X_i^\top (\Gamma^* + S^*)) + sd \cdot \epsilon_i,$$

onde $\{\epsilon_i\}_{i=1}^n$ é uma sequência iid de normais padrão.

```
In [7]: #Escreva o código aqui
def data_genXe(n ,d1, d2, B_true, sd):
    X = [np.random.normal(0, 1, (d1, d2)) for _ in range(n)]
    y = np.zeros(n)

    for i in range(n):
        epsilon_i = np.random.normal(0, 1)
        y[i] = np.trace(X[i].T @ (B_true)) + sd * epsilon_i

    return X, y.reshape(-1, 1)
```

```
In [8]: #Exemplo:
d_1=40
d_2=40
r=5
s=10
Gamma_mag=10
S_mag=10
sd = 1
n = 100

Gamma_true = data_genG(n,d_1,d_2,r,Gamma_mag)
S_true = data_genS(d_1,d_2,s,S_mag)
B_true = Gamma_true + S_true
X,y = data_genXe(n,d_1,d_2,B_true,sd)
```

```
In [9]: type(X), len(X), X[0].shape
```

```
Out[9]: (list, 100, (40, 40))
```

```
In [10]: type(y), len(y), y.shape
```

```
Out[10]: (numpy.ndarray, 100, (100, 1))
```

Exercício 5:

Construa uma função `df(Gamma,S)` que retorna o gradiente $\nabla f(\Gamma, S)$.

```
In [11]: #Escreva o código aqui
def df(Gamma, S, X, y):
    n = len(X)
    d1, d2 = Gamma.shape

    grad = np.zeros((d1, d2))

    for i in range(n):
        residual = y[i] - np.trace(X[i].T @ (Gamma + S))
        grad -= residual * X[i]

    return grad/n
```

```
In [12]: df(Gamma_true,S_true, X, y).shape, df(Gamma_true,S_true, X, y)
```

```
Out[12]: ((40, 40),
  array([[ 0.02663875,  0.029387 ,  0.02468258, ...,  0.02313545,
           0.14666829,  0.02782865],
        [-0.12411486, -0.06586148, -0.01630952, ...,  0.02317795,
          -0.00930994, -0.16260578],
        [ 0.00256909, -0.02995926, -0.06653458, ...,  0.05289242,
           0.11660331,  0.03563779],
        ...,
        [-0.16079459,  0.02570492, -0.08117275, ...,  0.03793658,
          -0.1290603 , -0.14656453],
        [ 0.17752151,  0.04505253, -0.02933661, ...,  0.1214401 ,
          -0.07851648,  0.13141602],
        [-0.11469431, -0.04126938, -0.02339187, ...,  0.02134529,
          -0.02431015, -0.1295621 ]]))
```

Exercício 6:

Vamos usar o método gradiente proximal com splitting para resolver o problema acima:

$$\begin{aligned}
W_{\Gamma,k+1} &:= \Gamma_k - \frac{1}{L} \nabla f(\Gamma_k, S_k), \\
W_{S,k+1} &:= S_k - \frac{1}{L} \nabla f(\Gamma_k, S_k), \\
\Gamma_{k+1} &:= P_N \left(W_{\Gamma,k+1}, \frac{\lambda}{L}, a \right), \\
S_{k+1} &:= P_1 \left(W_{S,k+1}, \frac{\tau}{L} \right).
\end{aligned}$$

Construa uma função `trace_reg_MD(Df, n,d_1,d_2,X,y,L,lambd,tau,a, Gamma0,S0,tfinal)` onde `Df` é uma função que retorna o gradiente, $1/L$ é o passo, `lambd` ($= \lambda$) e `tau` ($= \tau$) são os fatores de penalização, `Gamma0` e `S0` são os pontos iniciais, e a variável `tfinal` é o número de iterações. A função deve retornar a sequência $k \mapsto \|\nabla f(\Gamma_k, S_k)\|_F^2$ e os últimos iterados Γ_{tfinal} e S_{tfinal} . Use hyper-parâmetros:

$$\begin{aligned}
a &= \frac{\Gamma_{mag}}{\sqrt{n}}, \\
\lambda &= \text{sd} \sqrt{\frac{d_1 + d_2}{n}}, \\
\tau &= \text{sd} \sqrt{\frac{\log(d_1 d_2)}{n}} + \frac{\Gamma_{mag}}{\sqrt{n}}.
\end{aligned}$$

e passo A/L ajustando $A \geq 1$ com

$$L = A \cdot \lambda_{\max} \left(\frac{1}{n} \sum_{i=1}^n \text{vec}(X_i) \text{vec}^\top(X_i) \right),$$

onde $\text{vec}(X_i)$ representa a matrix $d_1 \times d_2$ X_i em forma vetorial de dimensão $d_1 d_2$.

In [13]:

```

#Escreva o código aqui
def trace_reg_MD(Df, n, d1, d2, X, y, L, lambd, tau, a, Gamma0, S0, tfinal):
    Gamma, S = Gamma0, S0
    grad_norms = list()

    for t in range(tfinal):
        grad = Df(Gamma, S, X, y)

        W_Gamma = Gamma - (1 / L) * grad
        W_S = S - (1 / L) * grad

        U, D, Vt = svd(W_Gamma, full_matrices=False)
        D_thresholded = csoft(D, lambd, a)
        Gamma = U @ np.diag(D_thresholded) @ Vt
        S = soft(W_S, tau / L)

        grad_norm = np.linalg.norm(grad) ** 2
        grad_norms.append(grad_norm)

    return grad_norms, Gamma, S

```

In [14]:

```

#Exemplo:
Gamma0 = np.ones((d_1,d_2))
S0 = np.ones((d_1,d_2))
tfinal = 100

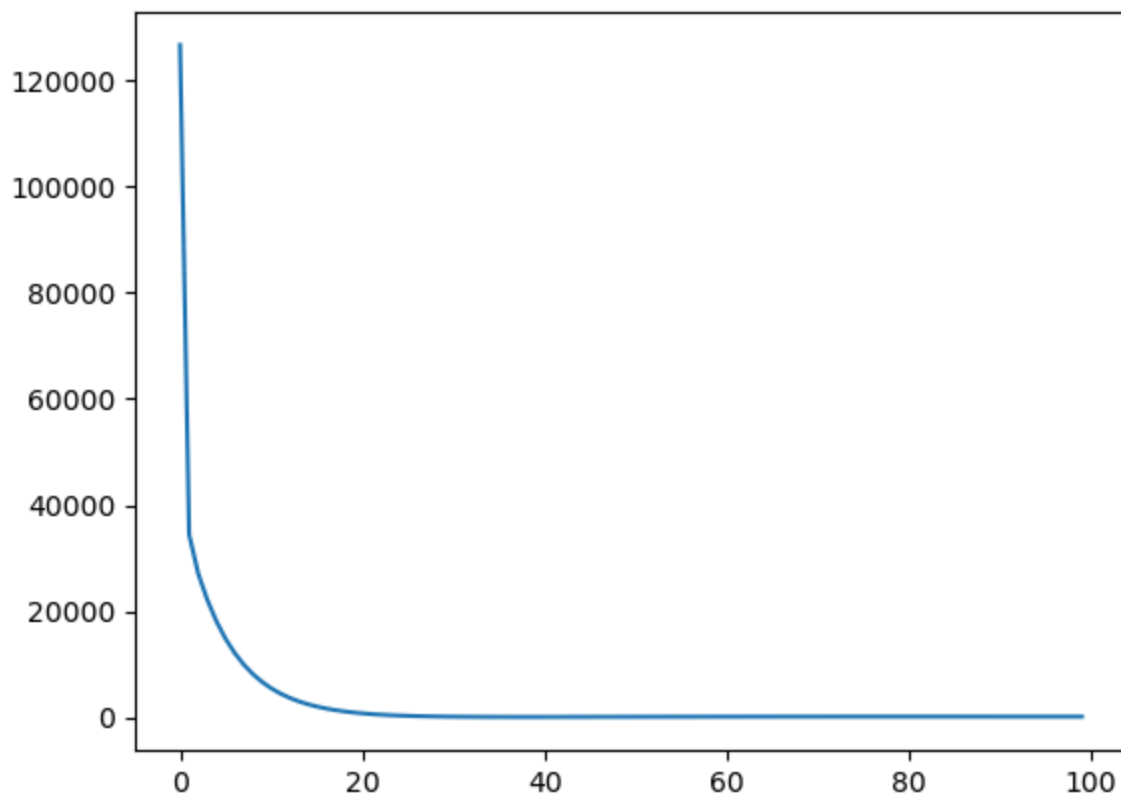
#Cálculo de L:
aux = []
for i in range(n):
    xi = np.asarray(X[i])
    aux.append(xi @ xi.T)
L = 4*np.max(la.eigvalsh(sum(aux)/n))

lambd = sd*np.sqrt((d_1+d_2)/n)
a = Gamma_mag/np.sqrt(n)
tau = sd*np.sqrt(np.log(d_1*d_2)/n) + Gamma_mag/np.sqrt(n)

f1 = trace_reg_MD(df, n,d_1,d_2,X,y,L,lambd,tau,a,Gamma0,S0,tfinal)
plt.plot(f1[0])

```

Out[14]: [`<matplotlib.lines.Line2D at 0x7f13941ee210>`]



Exercício 7:

O que acontece se A for muito pequeno?

Se o valor de A for muito pequeno, ele resultará em um passo $\frac{1}{L}$ grande demais no método de gradiente proximal. Com um passo excessivo, as atualizações de Γ e S podem se afastar muito dos pontos anteriores, resultando em instabilidade nas iterações e, eventualmente, fazendo com que o algoritmo não converja para uma solução ótima. Por fim, em vez de se moverem em direção ao mínimo, as iterações podem oscilar ou “pular” ao redor do ponto de convergência, sem nunca realmente estabilizar.

Exercício 8:

Agora, vamos usar o método gradiente proximal com *splitting acelerado*: iniciando de `[Gamma0,S0]= [Zg0,Zs0]` e $t_0 = 1$:

$$\begin{aligned} Z_{\Gamma,k+1} &:= P_N \left(\Gamma_k - (1/L) \nabla f(\Gamma_k, S_k), \frac{\lambda}{L}, a \right), \\ Z_{S,k+1} &:= P_1 \left(S_k - (1/L) \nabla f(\Gamma_k, S_k), \frac{\tau}{L} \right), \\ t_{k+1} &:= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ \Gamma_{k+1} &:= Z_{\Gamma,k+1} + \frac{t_k - 1}{t_{k+1}} (Z_{\Gamma,k+1} - Z_{\Gamma,k}), \\ S_{k+1} &:= Z_{S,k+1} + \frac{t_k - 1}{t_{k+1}} (Z_{S,k+1} - Z_{S,k}). \end{aligned}$$

Construa uma função `trace_reg_MD_acc(Df, n,d_1,d_2,X,y,L,lambd,tau,a,Gamma0,S0,tfinal)` onde `Df` é uma função que retorna o gradiente, $1/L$ é o passo, `lambd` ($= \lambda$) e `tau` ($= \tau$) são os fatores de penalização, `Gamma0` e `S0` são os pontos iniciais, e a variável `tfinal` é o número de iterações. A função deve retornar a sequência $k \mapsto \|\nabla f(\Gamma_k, S_k)\|_F^2$ e os últimos iterados Γ_{tfinal} e S_{tfinal} . Use hiperparâmetros:

$$\begin{aligned} a &= \frac{\Gamma_{mag}}{\sqrt{n}}, \\ \lambda &= \text{sd} \sqrt{\frac{d_1 + d_2}{n}}, \\ \tau &= \text{sd} \sqrt{\frac{\log(d_1 d_2)}{n}} + \frac{\Gamma_{mag}}{\sqrt{n}}. \end{aligned}$$

e passo A/L ajustando $A \geq 1$ com

$$L = A \cdot \lambda_{\max} \left(\frac{1}{n} \sum_{i=1}^n \text{vec}(X_i) \text{vec}^\top(X_i) \right).$$

```
In [15]: #Escreva o código aqui
def trace_reg_MD_acc(Df, n, d1, d2, X, y, L, lambd, tau, a, Gamma0, S0, tfinal):
    Gamma_k, S_k = Gamma0, S0
    Zg_k, Zs_k = Gamma0, S0
    t_k = 1
    gradient_norms = []

    for k in range(tfinal):
        grad = Df(Gamma_k, S_k, X, y)

        U, D, Vt = svd(Gamma_k - (1/L) * grad, full_matrices=False)
        D_thresholded = csoft(D, lambd / L, a)
        Zg_next = U @ np.diag(D_thresholded) @ Vt

        Zs_next = soft(S_k - (1/L) * grad, tau / L)

        t_next = (1 + np.sqrt(1 + 4 * t_k ** 2)) / 2

        Gamma_next = Zg_next + ((t_k - 1) / t_next) * (Zg_next - Zg_k)
        S_next = Zs_next + ((t_k - 1) / t_next) * (Zs_next - Zs_k)

        grad_norm = np.linalg.norm(grad) ** 2
        gradient_norms.append(grad_norm)

        Gamma_k, S_k = Gamma_next, S_next
        Zg_k, Zs_k = Zg_next, Zs_next
        t_k = t_next

    return gradient_norms, Gamma_k, S_k
```

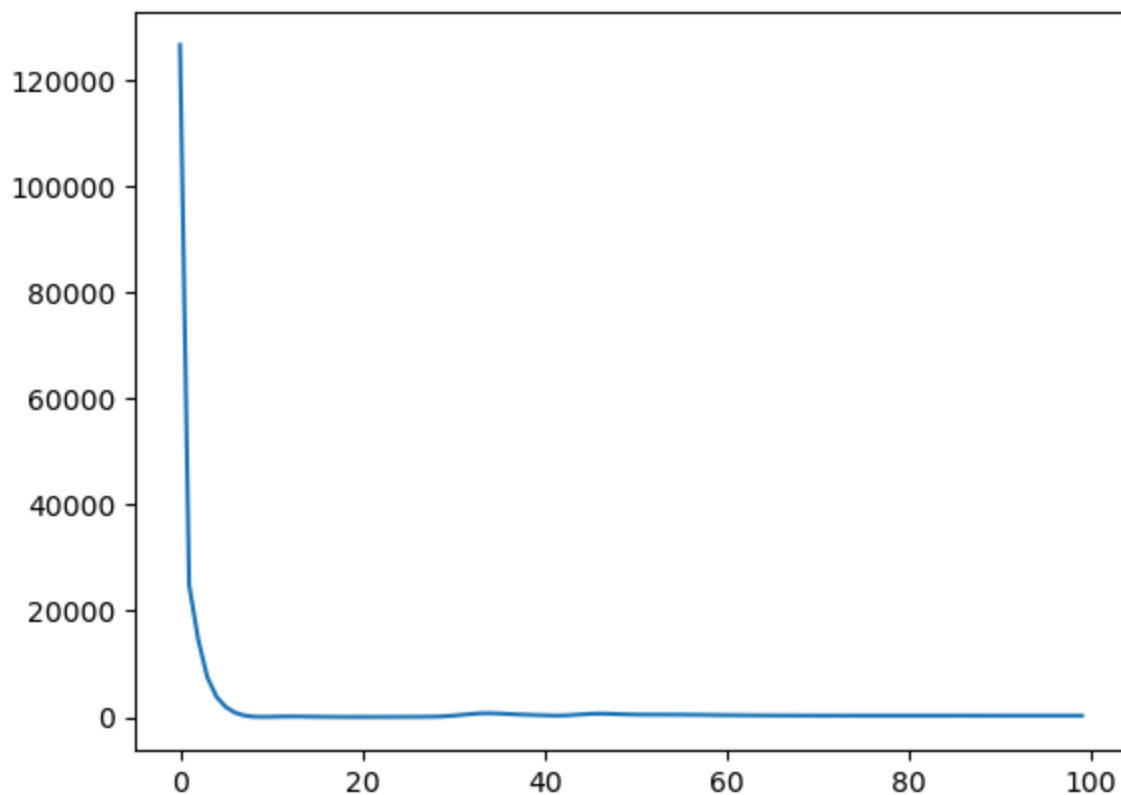
```
In [16]: Gamma0 = np.ones((d_1,d_2))
S0 = np.ones((d_1,d_2))
tfinal = 100

#Cálculo de L:
aux = []
for i in range(n):
    xi = np.asarray(X[i])
    aux.append(xi @ xi.T)
L = 4*np.max(la.eigvalsh(sum(aux)/n))

lambd = sd*np.sqrt((d_1+d_2)/n)
a = Gamma_mag/np.sqrt(n)
tau = sd*np.sqrt(np.log(d_1*d_2)/n) + Gamma_mag/np.sqrt(n)

f2 = trace_reg_MD_acc(df, n,d_1,d_2,X,y,L,lambd,tau,a,Gamma0,S0,tfinal)
plt.plot(f2[0])
```

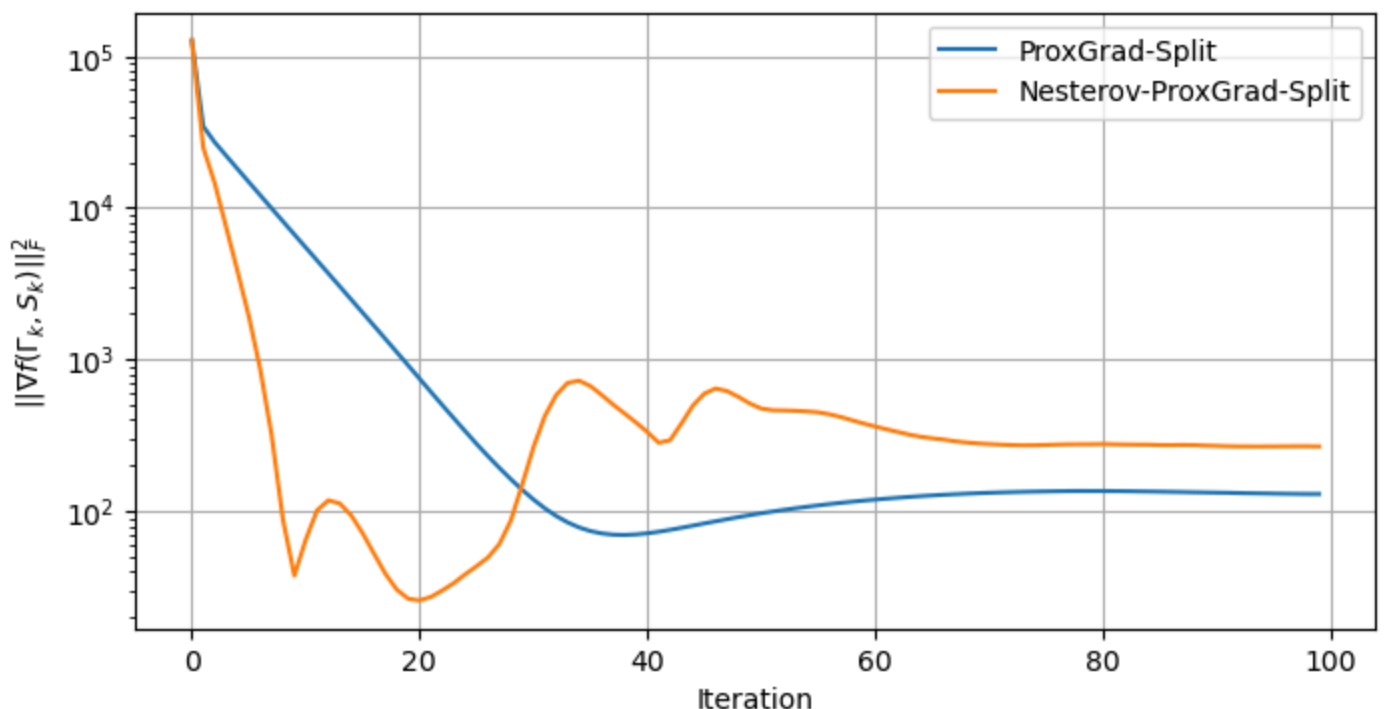
Out[16]: [`<matplotlib.lines.Line2D at 0x7f13927ebf10>`]



Exercício 9:

Implemente num mesmo gráfico os erros $k \mapsto \|\nabla f(\Gamma_k, S_k)\|_F^2$ de cada método em função no número de iterações.

```
In [17]: #Escreva o código aqui
plt.figure(figsize=(8, 4))
plt.plot(f1[0], label='ProxGrad-Split')
plt.plot(f2[0], label='Nesterov-ProxGrad-Split')
plt.yscale('log')
plt.grid(True)
plt.xlabel('Iteration')
plt.ylabel('$\|\nabla f(\Gamma_k, S_k)\|_F^2$')
plt.legend()
plt.show()
```



PCA robusta

O modelo anterior é *supervisionado*. Em *visão computacional*, um modelo relacionando mas não-supervisionado consiste em estimar cada uma das componentes $[\Gamma^*, S^*]$ tendo apenas uma medição ruidosa da soma $B^* = \Gamma^* + S^*$, por exemplo, ao separar de uma foto um carro (image esparsa) do ambiente (imagem de posto pequeno):

Neste caso, resolvemos:

$$\underset{[\Gamma,S] \in (\mathbb{R}^{d_1 \times d_2})^2}{\text{minimize}} \quad F(\Gamma, S) + \lambda \|\Gamma\|_N + \tau \|S\|_1,$$
$$\text{sujeito à} \quad \|\Gamma\|_\infty \leq a,$$

onde

$$F(\Gamma, S) = \frac{1}{2} \|Y - (\Gamma + S)\|_F^2,$$

Y é uma matriz de dados em $\mathbb{R} \times \mathbb{R}^{d_1 \times d_2}$, por exemplo, uma imagem ruidosa. Aplicações incluem *vigilância por vídeos, reconhecimento facial, topic modelling* (incluindo *latent semantic analysis* e *sentiment analysis*) e *recomendação de sistemas* sujeito a dados corruptos. Veja e.g. [Robust PCA?](#) e [Video](#).

Exercício 10:

Construa uma função `data_genYe(n,d_1,d_2,B_true,sd)` que toma n, d_1, d_2, B^* , e um número positivo `sd` e constrói a matriz de dados `Y` tal que

$$Y = B^* + \text{sd} \cdot \Xi,$$

onde Ξ é uma matriz $d_1 \times d_2$ cujas entradas são iid $N(0, 1)$. NOTA: Iremos usar as mesmas funções anteriores para gerar Γ^* e S^* .

```
In [18]: #Escreva o código aqui
def data_genYe(n, d1, d2, B_true, sd):
    Xi = np.random.randn(d1, d2)

    Y = B_true + sd * Xi

    return Y
```

```
In [19]: #Exemplo:
d_1=40
d_2=40
r=5
s=10
Gamma_mag=10
S_mag=10
sd = .1
n = 100

Gamma_true = data_genG(n,d_1,d_2,r,Gamma_mag)
S_true = data_genS(d_1,d_2,s,S_mag)
B_true = Gamma_true + S_true
Y = data_genYe(n,d_1,d_2,B_true,sd)
```

```
In [20]: Y.shape, Y
```

```
Out[20]: ((40, 40),
array([[10.13967737,  0.67049726, -0.47815347, ..., -0.1529142 ,
        -0.65545553,  0.28687972],
 [ 0.0602665 , -0.21075384, -0.06009046, ..., -0.0193544 ,
        0.30872886, -0.20946159],
 [ 0.31248775,  0.44651938,  0.17186607, ...,  0.17167906,
        0.09259373,  0.10967258],
 ...,
 [-0.05806322,  0.44734979, -0.17516871, ...,  0.041485 ,
        -0.31796687, -0.01205402],
 [ 0.09583256, -0.19366869, -0.05542686, ..., -0.15838352,
        0.32065859, -0.12578345],
 [-0.13611412,  0.34032916, -0.27543841, ...,  0.33003443,
        -0.42719854,  0.33173703]]))
```

Exercício 11:

Construa uma função `dF(Gamma,S)` que retorna o gradiente $\nabla F(\Gamma, S)$.

```
In [21]: #Escreva o código aqui
def dF(Gamma, S, X, Y):
```

```
error = Y - (Gamma + S)

df = -error

return df
```

Exercício 12:

Vamos usar o método gradiente proximal com splitting para resolver o problema acima. Ele tem exatamente a mesma expressão do Exercício 6 mas usando $\nabla F(\Gamma, S)$. Implemente o algoritmo com

$$a = \frac{\Gamma_{mag}}{\sqrt{n}},$$
$$\lambda = \text{sd} \sqrt{\frac{d_1 + d_2}{n}},$$
$$\tau = \text{sd} \sqrt{\frac{\log(d_1 d_2)}{n}} + \frac{\Gamma_{mag}}{\sqrt{n}}.$$

e passo $1/L$. PERGUNTA: Exiba um limite aproximado em que para L os métodos convergem e para um L próximo os métodos não convergem.

Em seguida, implemente o método gradiente proximal com splitting acelerado: é o mesmo que o exercício 8 mas usando $\nabla F(\Gamma, S)$.

Finalmente, plote num mesmo gráfico a sequência de erros $\|\nabla f(\Gamma_k, S_k)\|_F^2$ para os dois métodos.

```
In [22]: #Escreva o código aqui
a = Gamma_mag / np.sqrt(n)
lambd = sd * np.sqrt((d_1 + d_2) / n)
tau = sd * np.sqrt(np.log(d_1 * d_2) / n) + Gamma_mag / np.sqrt(n)

Gamma_init = np.random.randn(d_1, d_2)
S_init = np.random.randn(d_1, d_2)

f1_2 = trace_reg_MD(dF, n,d_1,d_2,X,Y,L,lambd,tau,a,Gamma0,S0,tfinal)
f2_2 = trace_reg_MD_acc(dF, n,d_1,d_2,X,Y,L,lambd,tau,a,Gamma0,S0,tfinal)

plt.figure(figsize=(10, 6))
plt.plot(f1_2[0], label='ProxGrad-Split')
plt.plot(f2_2[0], label='Nesterov-ProxGrad-Split')
plt.yscale("log")
plt.xlabel("Iterações")
plt.ylabel(r"$\| \nabla f(\Gamma_k, S_k) \|_F^2$")
plt.legend()
plt.show()
```

