Inversas generalizadas simétricas esparsas de matrizes simétricas

Ananias Sousa Machado **Marcia Fampa***
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, Brazil
E-mail: ananiasmachado@poli.ufrj.br, fampa@cos.ufrj.br,
Jon Lee†
University of Michigan
Ann Arbor, MI, USA
E-mail: jonxlee@umich.edu.

Resumo: As inversas generalizadas são úteis em uma ampla variedade de problemas de álgebra linear quando lidamos com matrizes que não possuem posto-linha ou posto-coluna completo. Para fins numéricos, é útil ter uma inversa generalizada que minimize uma norma e, ao mesmo tempo, seja esparsa. De fato, é natural escolher a norma 1 como um meio de induzir esparsidade. Quando a matriz dada é simétrica, podemos buscar uma inversa generalizada simétrica que minimize a norma 1. Apresentamos uma formulação para este problema e a utilizamos para desenvolver um algoritmo de ponto proximal Douglas-Rachford Splitting (DRS). Validamos nossa abordagem com experimentos numéricos.

Palavras-chave: propriedades de Moore-Penrose, inversa generalizada, norma 1 mínima, esparsidade, otimização linear, algoritmo Douglas–Rachford Splitting

1 Introdução

Quando uma matriz real $A \in \mathbb{R}^{m\times n}$ não é quadrada ou não é invertível, as inversas generalizadas de A são úteis para diversos problemas em álgebra matricial (ver [RM72]). Em particular, a pseudoinversa de Moore-Penrose (M-P) (descoberta independentemente por E.H. Moore e R. Penrose) pode ser usada para calcular a solução de vários problemas-chave em estatística, como a solução de mínimos quadrados de um sistema de equações lineares sobredeterminado e a solução de norma 2 mínima de um sistema de equações lineares subdeterminado. Geralmente, para estabilidade e eficiência numérica, pode ser útil ter uma inversa generalizada que otimize (ou otimize aproximadamente) diversas quantidades; por exemplo, norma 1 (vetorial) mínima, posto mínimo, esparsidade máxima (que é NP-difícil; veja [XFLP21, Prop. 1.3]). Tais matrizes geralmente não serão a pseudoinversa M-P. Em nosso cenário, assumimos que a matriz de entrada é simétrica e buscamos uma inversa generalizada simétrica que minimize a norma 1.

Se $A = U\Sigma V^\top$ é a decomposição em valores singulares reais de A (veja [GvL13], por exemplo), então a pseudoinversa M–P de A pode ser definida como $A^\dagger := V\Sigma^\dagger U^\top$, onde $\Sigma^\dagger$ tem a forma da transposta da matriz diagonal Σ, e é derivada de Σ tomando os recíprocos dos elementos não nulos (diagonais) de Σ (ou seja, os valores singulares não nulos de A). Podemos definir diferentes inversas generalizadas esparsas tratáveis, com base na seguinte caracterização fundamental da pseudoinversa M–P.

*Bolsista de Produtividade em Pesquisa do CNPq, bolsa 307167/2022-4
†Parcialmente financiado por AFOSR grant FA9550-22-1-0172

Teorema 1 ([Pen55]). *Para $A \in \mathbb{R}^{m\times n}$, a pseudoinversa M-P $A^\dagger$ é a única matriz $H \in \mathbb{R}^{n\times m}$ que satisfaz:*

$$AHA = A \tag{P1}$$
$$HAH = H \tag{P2}$$
$$(AH)^\intercal = AH \tag{P3}$$
$$(HA)^\intercal = HA \tag{P4}$$

Uma *inversa generalizada* de A é simplesmante qualquer matriz H que satisfaça P1.

A seguir, apresentaremos um resultado-chave para o nosso trabalho.

Teorema 2 ([BIG03], p. 208, Ex. 14 (sem demonstração); veja também [PFLX24] (com demosntração)). *Para uma matriz $A \in \mathbb{R}^{m\times n}$ de posto r, considere a decomposição completa em valores singulares $A =: U\Sigma V^\intercal$ com*

$$\Sigma =: \begin{bmatrix} \underset{r\times r}{D} & \underset{r\times(n-r)}{0} \\ \underset{(m-r)\times r}{0} & \underset{(m-r)\times(n-r)}{0} \end{bmatrix},$$

e D diagonal. Para $H \in \mathbb{R}^{n\times m}$, seja $\Gamma := V^\intercal HU$ (logo, $H = V\Gamma U^\intercal$), onde particionamos Γ em blocos como em

$$\Gamma =: \begin{bmatrix} \underset{r\times r}{X} & \underset{r\times(m-r)}{Y} \\ \underset{(n-r)\times r}{Z} & \underset{(n-r)\times(m-r)}{W} \end{bmatrix}.$$

Temos que

(i) P1 *é equivalente a $X = D^{-1}$.*
(ii) Se P1 *é satisfeita, então* P2 *é equivalente a $ZDY = W$.*
(iii) Se P1 *é satisfeita, então* P3 *é equivalente a $Y = 0$.*
(iv) Se P1 *é satisfeita, então* P4 *é equivalente a $Z = 0$.*

De fato, atribuindo a matriz nula a W, Y e Z no Teorema 2, podemos ver a relação simples entre $A^\dagger$ e a decomposição SVD de A.

Notação. Denotamos o i-ésimo vetor unitário padrão por e_i e a matriz identidade de ordem n por I_n. Para $p \in \{0\} \cup [1, \infty]$, denotamos a norma p vetorial ordinária por $\|\cdot\|_p$ (é claro que é apenas uma pseudonorma para $p = 0$). A vetorização de uma matriz $A \in \mathbb{R}^{m\times n}$ é denotada por $\operatorname{vec}(A) \in \mathbb{R}^{mn}$ e sua norma p vetorial é definida por $\|A\|_p := \|\operatorname{vec}(A)\|_p$. Para matrizes, $\operatorname{posto}(\cdot)$ denota o posto e $\|\cdot\|_F$ denota a norma de Frobenius. Denotamos o produto de Kronecker por $\otimes$ (usamos a identidade: $\operatorname{vec}(ABC) = (C^\intercal \otimes A)\operatorname{vec}(B)$). Denotamos por $\arg\min\{\cdot\}$ qualquer solução do problema de minimização associado.

2 Inversa generalizada simétrica esparsa de matriz simétrica

Consideramos o caso em que $A \in \mathbb{R}^{n\times n}$ é simétrica e buscamos uma inversa generalizada simétrica esparsa de A. Observamos que, se A é simétrica, então a pseudoinversa de Moore-Penrose de A é simétrica; portanto, sabemos que existem inversas generalizadas simétricas de A. Estamos interessados no seguinte problema, o qual busca induzir a esparsidade por meio da minimização da norma 1 vetorial.

$$\min_{H\in\mathbb{R}^{n\times n}} \{\|H\|_1 : P1 + (H^\intercal = H)\}. \tag{$P^1_{1,\mathrm{Sym}}$}$$

Uma alternativa seria simplesmente buscar uma inversa generalizada que minimize a norma 1, e então simetrizá-la, o que necessariamente ainda minimizaria a norma 1 devido à convexidade.

No entanto, com a simetrização, o posto poderia aumentar substancialmente e a esparsidade poderia diminuir substancialmente.

Podemos chegar a uma reformulação sutil, e útil, de $P^1_{1,\mathrm{Sym}}$ com a seguinte caracterização.

Teorema 3. *Seja $A \in \mathbb{R}^{n\times n}$ simétrica. H satisfaz P1 e $H^\top = H$ se e somente se H satisfaz $AHA + H = A + H^\top$.*

Proof. Seja $A = V\Sigma V^\top$ a decomposição completa em valores singulares de A. Consideremos $H = V\Gamma V^\top$ usando a notação do Teorema 2.

($\Rightarrow$) É fácil ver que se H satisfaz P1 e $H^\top = H$, então H satisfaz $AHA + H = A + H^\top$.

($\Leftarrow$) Para ver que H satisfaz P1 e $H^\top = H$ se H satisfaz $AHA + H = A + H^\top$, observe que

$$\begin{aligned}
AHA + H &= A + H^\top && \Leftrightarrow \\
V\Sigma V^\top V\Gamma V^\top V\Sigma V^\top + V\Gamma V^\top &= V\Sigma V^\top + V\Gamma^\top V^\top && \Leftrightarrow \\
\Sigma\Gamma\Sigma + \Gamma &= \Sigma + \Gamma^\top && \Leftrightarrow \\
\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X & Y \\ Z & W \end{bmatrix}\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} &= \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} X^\top & Z^\top \\ Y^\top & W^\top \end{bmatrix} && \Leftrightarrow \\
\begin{bmatrix} DXD + X & Y \\ Z & W \end{bmatrix} &= \begin{bmatrix} D + X^T & Z^T \\ Y^\top & W^\top \end{bmatrix}. &&
\end{aligned}$$

Portanto, $Z = Y^\top$, $W = W^\top$ e $DXD = D + X^\top$. De $DXD = D + X^\top$, temos

$$\begin{cases} d_i^2 x_{ii} + x_{ii} = d_i + x_{ii}\,, & \text{if } i = j; \\ d_i d_j x_{ij} + x_{ij} = x_{ji}\,, & \text{if } i \neq j. \end{cases}$$

Logo, temos que $d_i^2 x_{ii} + x_{ii} = d_i + x_{ii} \Leftrightarrow x_{ii} = \frac{1}{d_i}$. Somando as equações $d_i d_j x_{ij} + x_{ij} = x_{ji}$ e $d_i d_j x_{ji} + x_{ji} = x_{ij}$, para $i \neq j$, temos que $d_i d_j x_{ij} + x_{ij} + d_i d_j x_{ji} + x_{ji} = x_{ji} + x_{ij} \Leftrightarrow x_{ij} = -x_{ji}$. Portanto, $d_i d_j x_{ij} + x_{ij} = x_{ji} \Leftrightarrow d_i d_j x_{ij} + x_{ij} = -x_{ij} \Leftrightarrow (d_i d_j + 2)x_{ij} = 0$. Como $d_i, d_j > 0$, temos $x_{ij} = 0$, $\forall i \neq j$. Desta forma, $X = D^{-1}$. Pelo Teorema 2, H satisfaz P1. Como $Z = Y^\top$ e $W = W^\top$, H também satisfaz $H^\top = H$. □

Baseados no resultado acima, apresentamos a seguinte formulação equivalente para $P^1_{1,\mathrm{Sym}}$.

$$\min_{H\in\mathbb{R}^{n\times m}} \{\|H\|_1 : AHA + H = A + H^\top\}. \tag{$\tilde{P}^1_{1,\mathrm{Sym}}$}$$

3 Um algorithm DRS para $P^1_{1,\mathrm{Sym}}$

Desenvolvemos um algoritmo *Douglas-Rachford Splitting* (DRS) para resolver $P^1_{1,\mathrm{Sym}}$. O DRS é um método proximal que tem sido amplamente aplicado na literatura a esse tipo de problema com resultados promissores (veja, por exemplo, [DHP24, FZB20]). Ele resolve problemas de otimização convexa da forma

$$\min\ f(H) + g(H), \tag{1}$$

onde $f, g : \mathbb{R}^{n\times m} \to \mathbb{R} \cup \{\infty\}$ são convexas, fechadas e próprias, calculando iterativamente (see [DHP24], [BC17])

$$\begin{aligned}
H^{k+1/2} &= \mathrm{prox}_{\lambda f}(V^k), \\
V^{k+1/2} &= 2H^{k+1/2} - V^k, \\
H^{k+1} &= \mathrm{prox}_{\lambda g}(V^{k+1/2}), \\
V^{k+1} &= V^k + H^{k+1} - H^{k+1/2},
\end{aligned}$$

para $k = 0, 1, \ldots$, onde $\mathrm{prox}_{\lambda h}$ é dado por

$$\mathrm{prox}_{\lambda h}(V) = \arg\min_H \{h(H) + \frac{1}{2\lambda}\|H - V\|_F^2\}.$$

Estamos interessados em aplicar DRS para resolver $P^1_{1,\text{Sym}}$, ou, de forma mais geral, resolver um problema formulado como

$$\min\{\|H\|_1 : H \in \mathcal{C}\}, \tag{2}$$

onde $\mathcal{C}$ é um conjunto convexo. (2) pode ser escrito na forma de (1) tomando $f = \|\cdot\|_1$ e $g = \mathcal{I}_{\mathcal{C}}$ (a função característica do conjunto convexo $\mathcal{C}$). Nesse caso, temos de [PB14] que os respectivos operadores proximais são dados por $\text{prox}_{\lambda f}(X) = S_\lambda(X)$ e $\text{prox}_{\lambda g}(X) = \Pi_{\mathcal{C}}(X)$, onde $S_\lambda(X)$ é o operador de *soft thresholding* elemento a elemento, definido por

$$S_\lambda(x) = \begin{cases} x - \lambda, & a > \lambda; \\ 0, & |x| \leq \lambda; \\ x + \lambda, & a < -\lambda, \end{cases}$$

e $\Pi_{\mathcal{C}}$ é a projeção sobre o conjunto convexo $\mathcal{C}$.

A seguir, abordamos mais detalhes sobre a aplicação do DRS ao problema $P^1_{1,\text{Sym}}$, começando com a projeção sobre o conjunto específico $\mathcal{C}$ para o problema. Em muitos casos, como o nosso, em que $\mathcal{C}$ é afim, a expressão da projeção possui uma solução analítica, que desenvolvemos a seguir.

3.1 Projeção sobre $\mathcal{C}$

O conjunto de todas as inversas generalizadas simétricas de uma dada matriz simétrica A pode ser expresso como o espaço afim $\mathcal{C} = \{H : AHA = A, H^\intercal = H\}$. Abordaremos agora o cálculo de uma projeção sobre este conjunto.

Proposição 4. *Se* $\mathcal{C} = \{H : AHA = A, H^\intercal = H\}$, *então*

$$\Pi_{\mathcal{C}}(V) = \tfrac{1}{2}(V + V^\intercal) - \tfrac{1}{2}AA^\dagger(V + V^\intercal)A^\dagger A + A^\dagger.$$

Proof. Seja $V \in \mathbb{R}^{n\times m}$ uma matriz dada. Queremos encontrar a solução analítica de

$$\arg\min_H \{\|H - V\|_F^2 : AHA = A, H^\intercal = H\}. \tag{Proj$_{1\text{Sym}}$}$$

O função lagrangiana para Proj$_{1\text{Sym}}$ é dada por

$$\mathcal{L}(H, \Lambda, \Gamma) = \|H - V\|_F^2 + \langle \Lambda, AHA - A\rangle + \langle \Gamma, H - H^\intercal\rangle. \tag{3}$$

Como Proj$_{1\text{Sym}}$ é convexo, suas soluções primal e dual (H, Λ, Γ) são tais que $\nabla_H \mathcal{L}(H, \Lambda, \Gamma) = 0$, ou equivalentemente, tais que

$$\begin{aligned}
H &= V - \tfrac{1}{2}(A\Lambda A + \Gamma - \Gamma^\intercal) && \Rightarrow \quad (4)\\
AHA = A &= AVA - \tfrac{1}{2}A(A\Lambda A + \Gamma - \Gamma^\intercal)A && \Leftrightarrow \\
A^2\Lambda A^2 &= 2(AVA - A) - A(\Gamma - \Gamma^\intercal)A && \Rightarrow \\
A^{\dagger 2}A^2\Lambda A^2 A^{\dagger 2} &= 2A^{\dagger 2}(AVA - A)A^{\dagger 2} - A^{\dagger 2}A(\Gamma - \Gamma^\intercal)AA^{\dagger 2} && \Leftrightarrow \\
A^\dagger A\Lambda AA^\dagger &= 2(A^\dagger VA^\dagger - A^{\dagger 3}) - A^\dagger(\Gamma - \Gamma^\intercal)A^\dagger && \Leftrightarrow \\
((AA^\dagger) \otimes (A^\dagger A))\,\text{vec}(\Lambda) &= \text{vec}(2(A^\dagger VA^\dagger - A^{\dagger 3}) - A^\dagger(\Gamma - \Gamma^\intercal)A^\dagger).
\end{aligned}$$

Como $(AA^\dagger) \otimes (A^\dagger A)$ é uma projeção ortogonal, $\Lambda = 2(A^\dagger VA^\dagger - A^{\dagger 3}) - A^\dagger(\Gamma - \Gamma^\intercal)A^\dagger$ é uma solução para a última equação acima. Substituindo este valor de Λ em (4), temos

$$\begin{aligned}
H &= V - A(A^\dagger VA^\dagger - A^{\dagger 3})A + \tfrac{1}{2}AA^\dagger(\Gamma - \Gamma^\intercal)A^\dagger A - \tfrac{1}{2}(\Gamma - \Gamma^\intercal) && \Leftrightarrow \\
H &= V - AA^\dagger VA^\dagger A + A^\dagger + \tfrac{1}{2}AA^\dagger(\Gamma - \Gamma^\intercal)A^\dagger A - \tfrac{1}{2}(\Gamma - \Gamma^\intercal) && \Leftrightarrow \\
H^\intercal &= V^\intercal - AA^\dagger V^\intercal A^\dagger A + A^\dagger + \tfrac{1}{2}AA^\dagger(\Gamma^\intercal - \Gamma)A^\dagger A - \tfrac{1}{2}(\Gamma^\intercal - \Gamma) && \Rightarrow \\
H - H^\intercal &= V - V^\intercal - AA^\dagger(V - V^\intercal)A^\dagger A + AA^\dagger(\Gamma - \Gamma^\intercal)A^\dagger A - (\Gamma - \Gamma^\intercal) && \Rightarrow \\
0 &= V - V^\intercal - AA^\dagger(V - V^\intercal)A^\dagger A + AA^\dagger(\Gamma - \Gamma^\intercal)A^\dagger A - (\Gamma - \Gamma^\intercal) && \Leftrightarrow
\end{aligned}$$

$$(\Gamma - \Gamma^\intercal) - AA^\dagger(\Gamma - \Gamma^\intercal)A^\dagger A = V - V^\intercal - AA^\dagger(V - V^\intercal)A^\dagger A.$$

Assim, $\Gamma = V$ resolve a última equação acima. Substituindo agora os valores de Λ e Γ em (4), temos

$$\begin{aligned} H &= V - AA^\dagger VA^\dagger A + A^\dagger + \tfrac{1}{2}AA^\dagger(V - V^\intercal)A^\dagger A - \tfrac{1}{2}(V - V^\intercal) \\ &= \tfrac{1}{2}(V + V^\intercal) - \tfrac{1}{2}AA^\dagger(V + V^\intercal)A^\dagger A + A^\dagger, \end{aligned}$$

que é a expressão para a projeção. □

3.2 Ponto inicial e critério de parada

Consideramos como ponto inicial $V^0 := A^\dagger$, que é uma solução viável para o nosso problema e não requer custo computacional adicional, pois já é utilizada no cálculo da projeção $\Pi_{\mathcal{C}}(\cdot)$.

Para o critério de parada, assumimos que ϵ^{abs} e ϵ^{rel} são tolerâncias positivas dadas e calculamos $\tau^k := H^{k+1} - H^{k+1/2} = V^{k+1} - V^k$ em cada iteração k. Interrompemos o algoritmo se $||\tau^k||_F \leq \epsilon^{\text{abs}} + \epsilon^{\text{rel}}||V^1 - V^0||_F$, $k > 0$, que busca tanto uma diferença pequena entre as duas últimas iterações quanto uma diferença pequena relativa à diferença entre as duas iterações iniciais. Note que, de [BC17], temos que se a sequência V^k $(k = 0, 1, 2, \ldots)$ converge para um ponto fixo de $F(V) := V + \Pi_{\mathcal{C}}(2S_\lambda(V) - V) - S_\lambda(V)$, então ambas as sequências $H^{k+1/2}$ $(k = 0, 1, 2, \ldots)$ e H^{k+1} $(k = 0, 1, 2, \ldots)$ convergem para H^*, uma solução do problema (1). Finalmente, observamos que sempre obtemos uma solução primal viável do algoritmo DRS, tomando como saída a matriz H^{k+1} calculada na última iteração k, já que H^{k+1} é obtida a partir de uma projeção no conjunto viável do problema abordado.

4 Limite na esparsidade dos pontos extremos do problema linear

O problema que resolvemos, buscando inversas generalizadas simétricas esparsas, induz esparsidade com a norma 1 do vetor e, portanto, pode ser reformulado como um problema de otimização linear. É interessante caracterizar os pontos extremos do conjunto viável desse problema, pois essa caracterização pode nos fornecer um limite superior para o número de elementos não nulos nas soluções viáveis básicas para esse problema e, portanto, um limite superior para o número de elementos não nulos na inversa generalizada simétrica mais esparsa. Gostaríamos de enfatizar que um ponto extremo arbitrário (esparso) pode ter uma norma 1 vetorial grande e, por razões numéricas, isso é indesejável. Para comparações em nossos experimentos numéricos, apresentamos a seguir um resultado de [XFLP21] que caracteriza esses pontos extremos.

Proposição 5. [XFLP21, Prop. 2.1] *Suponha que $A \in \mathbb{R}^{n\times n}$ é simétrica e tem posto r. Então, as soluções extremas do problema de programação linear associado a $P^1_{1,\text{Sym}}$ têm no máximo $r^2 + r$ elementos não nulos. Além disso, este limite é alcançado para $n - 2 \geq r \geq 3$.*

5 Resultados numéricos

Apresentamos resultados numéricos usando o solver Gurobi v12.0.1 para resolver $P^1_{1,\text{Sym}}$ e $\tilde{P}^1_{1,\text{Sym}}$ e o algoritmo DRS apresentado na Seção 3. Usamos a linguagem de programação Julia v1.11.3 (usando VSCodium v1.97.1), e executamos os experimentos na "zebra", uma máquina de 16 núcleos (executando Windows Server 2019 Standard) com processadores Intel Xeon E5-2667 v4 rodando a 3,20 GHz com 8 núcleos cada, e 128 GB de memória. Geramos aleatoriamente 140 matrizes densas $B \in \mathbb{R}^{n\times n}$ com posto r (5 para cada configuração dada por (n, r)), com a função *sprand* do MATLAB, seguindo a maneira descrita em [FLPX21]. Para nossas instâncias de teste, calculamos então as matrizes simétricas $A = B^\intercal B$. Os parâmetros que usamos foram $n = 100, 200, \ldots, 500, 1000, 2000, \ldots, 5000$, $r = 0,25n$. Em todos os experimentos, o limite de tempo usado foi de 7200 segundos e as estatísticas apresentadas são a média calculada para cada conjunto de 5 instâncias da mesma configuração. Os parâmetros usados para o Gurobi são: *Barrier convergence tolerance*: 10^{-5}, *Feasibility tolerance*: 10^{-5} e *Optimality tolerance*: 10^{-5}; e para DRS_{FP} são: $\epsilon^{\text{abs}} = 10^{-5}$, $\epsilon^{\text{rel}} = 10^{-3}$ e $\lambda = 10^{-2}$.

Resolvemos $P^1_{1,\mathrm{Sym}}$ e $\tilde{P}^1_{1,\mathrm{Sym}}$ com o Gurobi. Só conseguimos resolver as instâncias de dimensão $n = 100$ no tempo limite, obtendo os resultados médios mostrados na Tabela 1. Mostramos a razão entre a norma 0 de H e o limite superior r^2+r no número de elementos não nulos de soluções extremas da reformulação de programação linear de $P^1_{1,\mathrm{Sym}}$ (ver Proposição 5). Comparamos a solução H obtida com o Gurobi com a pseudoinversa de Moore-Penrose de A, mostrando as razões $\|H\|/\|A^\dagger\|$ para a norma 0 e a norma 1. Também mostramos a razão entre o posto de H e r.

					$P^1_{1,\mathrm{Sym}}$	$\tilde{P}^1_{1,\mathrm{Sym}}$
n	$\frac{\|H\|_0}{r^2+r}$	$\frac{\|H\|_0}{\|A^\dagger\|_0}$	$\frac{\|H\|_1}{\|A^\dagger\|_1}$	$\frac{\mathrm{r}(H)}{r}$	tempo (segundos)	tempo (segundos)
100	0.913	0.064	0.447	2.456	403.15	560.85

Table 1: Resultados para Gurobi para $P^1_{1,\mathrm{Sym}}$ and $\tilde{P}^1_{1,\mathrm{Sym}}$ ($r = 0.25n$)

Estatísticas semelhantes são apresentadas na Tabela 2 para os resultados obtidos com o algoritmo DRS. Comparando os resultados obtidos pelo DRS para $n = 100$ com os resultados obtidos com Gurobi, observamos que as soluções do DRS têm a mesma norma 1, mas são menos esparsas, com norma 0 cerca de 28% maior. No entanto, o tempo computacional necessário para o Gurobi resolver este problema o torna não competitivo com o DRS. Notamos ainda que todas as soluções do DRS são muito mais esparsas do que $A^\dagger$, demonstrando a eficácia da indução de esparsidade com a norma 1 em nossos modelos. Ressaltamos que conseguimos resolver todas as instâncias até $n = 3500$ com o DRS dentro do tempo limite, enquanto com o Gurobi conseguimos resolver apenas as instâncias menores. Além disso, conseguimos resolver as menores instâncias em uma ordem de magnitude menor em tempo do que o Gurobi. Finalmente, observamos que, em média, a norma 0 das soluções obtidas por Gurobi e DRS fica próxima de $r^2 + r$, limite superior para o número de elementos não nulos nas soluções extremas dos programas lineares associados. Mesmo sem explorar os pontos extremos na resolução do problema, o DRS fica sempre próximo do limite (em média não mais que 30% acima) e abaixo dele nas instâncias maiores. Por fim, notamos que as soluções esparsas obtidas têm posto maior que o posto r de $A^\dagger$; portanto, podemos considerar o aumento no posto como uma contrapartida para alcançar maior esparsidade com nossa abordagem. Concluímos que nosso modelo para induzir esparsidade na inversa generalizada simétrica foi eficaz em nosso experimento, assim como o algoritmo DRS proposto para resolver o problema.

n	$\frac{\|H\|_0}{r^2+r}$	$\frac{\|H\|_0}{\|A^\dagger\|_0}$	$\frac{\|H\|_1}{\|A^\dagger\|_1}$	$\frac{\mathrm{r}(H)}{r}$	tempo (segundos)
100	1.167	0.082	0.447	2.62	10.90
200	1.266	0.085	0.384	2.98	16.62
300	1.184	0.081	0.383	2.91	17.31
400	1.114	0.076	0.374	3.10	60.13
500	1.148	0.078	0.359	3.21	109.14
1000	1.048	0.073	0.331	3.36	411.77
1500	1.009	0.071	0.323	3.50	841.76
2000	0.988	0.070	0.315	3.53	1489.70
2500	0.982	0.070	0.310	3.59	2639.73
3000	0.967	0.070	0.303	3.60	3574.90
3500	0.962	0.070	0.297	3.66	4780.89

Table 2: Resultados para DRS para $P^1_{1,\mathrm{Sym}}$ ($r = 0.25n$)

References

[BC17] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer Science & Business Media, 2017. https://doi.org/10.1007/978-3-319-48311-5.

[BIG03] Adi Ben-Israel and Thomas N.E. Greville. *Generalized Inverses: Theory and Applications*. Springer, 2nd edition, 2003. https://doi.org/10.1007/b97366.

[DHP24] Charles Dossal, Samuel Hurault, and Nicolas Papadakis. Optimization with first order algorithms, 2024. https://arxiv.org/abs/2410.19506.

[FLPX21] Marcia Fampa, Jon Lee, Gabriel Ponte, and Luze Xu. Experimental analysis of local searches for sparse reflexive generalized inverses. *Journal of Global Optimization*, 81:1057–1093, 2021. https://doi.org/10.1007/s10898-021-01087-y.

[FZB20] Anqi Fu, Junzi Zhang, and Stephen Boyd. Anderson accelerated Douglas-Rachford splitting. *SIAM Journal on Scientific Computing*, 42(6):A3560–A3583, 2020. http://dx.doi.org/10.1137/19M1290097.

[GvL13] Gene H. Golub and Charles F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 4th edition, 2013. https://epubs.siam.org/doi/abs/10.1137/1.9781421407944.

[PB14] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014. https://doi.org/10.1561/2400000003.

[Pen55] Roger Penrose. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51:406–413, 1955. https://doi.org/10.1017/S0305004100030401.

[PFLX24] Gabriel Ponte, Marcia Fampa, Jon Lee, and Luze Xu. On computing sparse generalized inverses. *Operations Research Letters*, 52:107058, 2024. https://doi.org/10.1016/j.orl.2023.107058.

[RM72] Calyampudi R. Rao and Sujit Kumar Mitra. Generalized inverse of a matrix and its applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. Volume I: Theory of Statistics (Lucien M. Le Cam, Jerzy Neyman, Elizabeth L. Scott, eds., University of California Press)*, pages 601–620, 1972. https://doi.org/10.1525/9780520325883-032.

[XFLP21] Luze Xu, Marcia Fampa, Jon Lee, and Gabriel Ponte. Approximate 1-norm minimization and minimum-rank structured sparsity for various generalized inverses via local search. *SIAM Journal on Optimization*, 31(3):1722–1747, 2021. https://doi.org/10.1137/19M1281514.