

Matrizes positivas definidas, semidefinidas, etc.

Amit Bhaya,
Programa de Engenharia Elétrica
COPPE/UFRJ
Universidade Federal do Rio de Janeiro
amit@nacad.ufrj.br
<http://www.nacad.ufrj.br/~amit>

Funções quadráticas positivas definidas

Uma função quadrática $f(x, y) = ax^2 + 2bxy + cy^2$ é chamada **positiva definida (p.d.)** se $\forall x, y \in \mathbb{R}, (x, y) \neq (0, 0), f(x, y) > 0$.

Condições necessárias e suficientes para que uma função seja p.d. Completamos o quadrado:

$$f = a \left(x + \frac{b}{a}y \right)^2 + \left(c - \frac{b^2}{a} \right) y^2$$

percebendo que $f > 0$ se e somente se $a > 0$ e $ac - b^2 > 0$.

Conclusão: Uma função $F(x, y)$ de duas variáveis $x, y \in \mathbb{R}$ atinge um mínimo (não-degenerado) no ponto $x = y = 0$ se e somente se suas primeiras derivadas parciais forem nulas e as segundas satisfazem

$$\frac{\partial^2 F}{\partial x^2} \Big|_{(0,0)} > 0, \quad \left[\frac{\partial^2 F}{\partial x^2} \Big|_{(0,0)} \quad \frac{\partial^2 F}{\partial y^2} \Big|_{(0,0)} \right] > \left[\frac{\partial^2 F}{\partial x \partial y} \Big|_{(0,0)} \right]^2.$$

Formas quadráticas positivas definidas

Dada uma matriz \mathbf{A} simétrica (hermitiana, resp.), a **forma quadrática** é definida como $Q_{\mathbf{A}}(\mathbf{x}) := \mathbf{x}^T \mathbf{A} \mathbf{x}$, para $\mathbf{x} \in \mathbb{R}^n$ (resp. \mathbb{C}^n). Se $\forall \mathbf{x} \neq 0, Q_{\mathbf{A}}(\mathbf{x}) > 0$, então a forma quadrática é chamada **positiva definida (p.d.)**.

Para $\mathbf{A} = (a_{ij})$, $Q_{\mathbf{A}}(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$. Isto é, a forma por extenso mostra explicitamente que a forma quadrática contém somente termos de grau 2 (do tipo x_i^2 ou $x_i x_j$), justificando o nome forma *quadrática*.

Sendo $\mathbf{H} = (h_{ij})$ a **hessiana** (matriz de segundas derivadas parciais de uma função F , i.e., $h_{ij} := \partial^2 F / \partial x_i \partial x_j$), podemos dizer que F possui um mínimo quando a forma quadrática $Q_{\mathbf{H}}(\mathbf{x}) = \mathbf{x}^T \mathbf{H} \mathbf{x}$ é p.d.

A expansão de Taylor da função F em torno de $\mathbf{0}$ é:

$$F(\mathbf{x}) = F(\mathbf{0}) + \mathbf{x}^T \text{grad } F + \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x} + \text{termos de ordem } > 3,$$

onde $\text{grad } \mathbf{F} := (\partial F / \partial x_1, \dots, \partial F / \partial x_n)$, (o gradiente de \mathbf{F}) se anula no mínimo.

Testes para p.d.

Uma matriz é chamada **positiva definida (p.d.)** se a forma quadrática associada a \mathbf{A} , $Q_{\mathbf{A}}(\mathbf{x}) > 0, \forall \mathbf{x} \neq \mathbf{0}$.

Testes para \mathbf{A} p.d.: Os seguintes testes são condições necessárias e suficientes *equivalentes* para que uma matriz $\mathbf{A} = \mathbf{A}^T$ seja positiva definida.

1. $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \forall \mathbf{x} \neq \mathbf{0}$.
2. Todos os autovalores de \mathbf{A} são positivos.
3. Todas as matrizes líderes¹ possuem determinantes positivos.
4. Todos os pivôs são positivos (e não é preciso, teoricamente, fazer trocas de linhas na eliminação gaussiana em \mathbf{A}).
5. $\exists \mathbf{R}$ com colunas l.i. tal que $\mathbf{A} = \mathbf{R}^T \mathbf{R}$.

¹formadas pelos cantos superiores à esquerda, sempre escolhendo as k primeiras linhas e colunas, $k = 1, 2, \dots$)

Matrizes p.d. e elipsóides

Seja $\mathbf{A} = \mathbf{A}^T$ positiva definida. Sabemos (pelo teorema espectral) que $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, com os autovalores $\lambda_i > 0$. A matriz ortogonal (de rotação) simplifica $\mathbf{x}^T \mathbf{A} \mathbf{x} = 1$:

$$\mathbf{x}^T \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \mathbf{x} = 1 \quad \text{ou} \quad \mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = 1 \quad \text{ou} \quad \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 = 1.$$

Esta é a equação de um elipsóide com semi-eixos apontando (no espaço x original) nas direções dos autovetores e tendo comprimentos $1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_n}$.

Matrizes semidefinidas

Uma matriz é chamada **positiva semi-definida (p.s.d.)** se a forma quadrática associada a \mathbf{A} , $Q_{\mathbf{A}}(\mathbf{x}) \geq 0, \forall \mathbf{x} \neq \mathbf{0}$.

Testes para \mathbf{A} p.s.d.: Os seguintes testes são condições necessárias e suficientes *equivalentes* para que uma matriz $\mathbf{A} = \mathbf{A}^T$ seja positiva semi-definida.

1. $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0, \forall \mathbf{x} \neq \mathbf{0}$.
2. Todos os autovalores de \mathbf{A} são não-negativos (≥ 0).
3. **Todas as submatrizes principais²** possuem determinantes não-negativos.
4. Todos os pivôs são não-negativos.
5. $\exists \mathbf{R}$ tal que $\mathbf{A} = \mathbf{R}^T \mathbf{R}$ (cols. de \mathbf{R} podem ser l.d.!).

²formadas pela escolha de *qualquer conjunto* de k linhas e as mesmas colunas, $k = 1, 2, \dots$)

Congruências e Inércia

Operação de mudança de variáveis em forma quadrática leva a idéia de congruência. Seja $Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, e $\mathbf{x} = \mathbf{C} \mathbf{y}$ (mudança de variáveis). Então definimos uma forma quadrática no vetor \mathbf{y} , $\mathbf{y}^T \mathbf{C}^T \mathbf{A} \mathbf{C} \mathbf{y} =: Q_{\mathbf{B}}(\mathbf{y})$, onde $\mathbf{B} := \mathbf{C}^T \mathbf{A} \mathbf{C}$. A matriz \mathbf{B} é chamada **congruente** a \mathbf{A} , e a transformação $\mathbf{A} \rightarrow \mathbf{C}^T \mathbf{A} \mathbf{C}$, \mathbf{C} não-singular é denominada **transformação de congruência**.

Lei de inércia de Sylvester: A matriz $\mathbf{C}^T \mathbf{A} \mathbf{C}$ possui o mesmo número de autovalores positivos, negativos e zero que \mathbf{A} . Ou seja, congruências conservam os sinais dos autovalores, não as localizações.

Esboço da prova da lei de inércia

Vamos provar o caso onde \mathbf{A} é não-singular ($\rightarrow \mathbf{C}^T \mathbf{A} \mathbf{C}$ é não-singular também), e não precisamos considerar autovalores nulos.

Prova: Suponha que \mathbf{C} pode ser transformada continuamente ($\mathbf{C}(t)$ varia continuamente no intervalo $[0, 1]$) em uma \mathbf{Q} ortogonal: i.e., $\mathbf{C}(0) = \mathbf{C}$ e $\mathbf{C}(1) = \mathbf{Q}$. Então os autovalores de $\mathbf{C}(t)^T \mathbf{A} \mathbf{C}(t)$ também mudam continuamente, passando do conjunto dos AV de $\mathbf{C}^T \mathbf{A} \mathbf{C}$ ($t = 0$) para o conjunto dos AV de $\mathbf{Q}^T \mathbf{A} \mathbf{Q}$ ($t = 1$). Como $\mathbf{C}(t)$ nunca é singular, *nenhum desses AV (são reais!) pode tocar/cruzar a origem*. Portanto, o número de AV à esquerda da origem, e o número à esquerda, para $\mathbf{C}^T \mathbf{A} \mathbf{C}$ e $\mathbf{Q}^T \mathbf{A} \mathbf{Q}$ têm que ser iguais. E estes números são os mesmos dos de \mathbf{A} que possui os mesmos AV que a matriz similar $\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} = \mathbf{Q}^T \mathbf{A} \mathbf{Q}$.

Para terminar a prova, temos que mostrar a transformação contínua. Aplique Gram-Schmidt às colunas de \mathbf{C} . Então $\mathbf{C} = \mathbf{Q} \mathbf{R}$, e a transformação contínua é: $\mathbf{C}(t) = t \mathbf{Q} + (1 - t) \mathbf{Q} \mathbf{R}$. A família de matrizes $\mathbf{C}(t)$ pode ser escrita como $\mathbf{C}(t) = \mathbf{Q}(t \mathbf{I} + (1 - t) \mathbf{R})$, portanto também passa pelo G-S, com fatores \mathbf{Q} e $t \mathbf{I} + (1 - t) \mathbf{R}$, ambos inversíveis. (porque?)

Aplicação da lei de inércia

Para qualquer matriz simétrica, *os sinais dos pivôs coincidem com os sinais dos autovalores*. A matriz Λ dos autovalores e a matriz dos pivôs possui o mesmo número de elementos positivos, negativos e zero.

Prova. $\mathbf{A} = \mathbf{LDL}^T$!

Localização dos AV através de deslocamentos:
Exemplo:

$$\mathbf{A} = \begin{bmatrix} 3 & 3 & 0 \\ 3 & 10 & 7 \\ 0 & 7 & 8 \end{bmatrix}, \quad \text{e} \quad \mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 8 & 7 \\ 0 & 7 & 6 \end{bmatrix}.$$

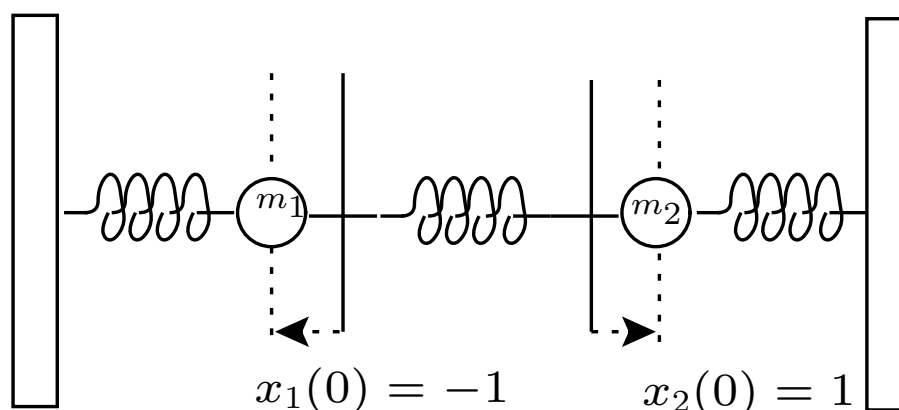
Calculando: os pivôs de \mathbf{A} são todos positivos, $\mathbf{A} - 2\mathbf{I}$ possui um pivô negativo. Conclusão: AV de \mathbf{A} positivos, e *um deles é menor do que 2* (porque)?

Próximo passo: Deslocamento $\mathbf{A} - \mathbf{I}$ possui pivô negativo ($\rightarrow AV \leq 1$).

$\mathbf{A} \rightarrow$ tridiagonal \rightarrow pivôs in $2n$ passos (não $\frac{1}{6}n^3!$), eliminação, pivôs, deslocamentos, biseção ...: *método de Givens*

Problema de autovalor generalizado: motivação

Sistema oscilatório com duas massas diferentes e molas.



Equações (usando Lei de Newton: $F = m\ddot{x}$): Deslocamentos das massas x_1 e x_2 , força das molas; $F_1 = -2x_1 + x_2$, $F_2 = x_1 - 2x_2$. Portanto:

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \mathbf{x}$$

Se as massas forem iguais, chegaríamos a $\ddot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ (familiar). Agora, tem-se $\mathbf{M}\ddot{\mathbf{x}} = \mathbf{A}\mathbf{x}$. O problema do **AV generalizado** surge quando procuramos soluções oscilatórias, i.e., $e^{i\omega t}$:

$$\mathbf{M}\ddot{\mathbf{x}} = \mathbf{A}\mathbf{x} \rightarrow \mathbf{M}(i\omega)^2 e^{i\omega t} \mathbf{x} = \mathbf{A}e^{i\omega t} \mathbf{x}.$$

Autovalores generalizados (continuação)

$$\mathbf{Ax} = \lambda \mathbf{Mx}, \quad \text{ou} \quad \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \mathbf{x} = \lambda \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \mathbf{x}.$$

Solução se $\lambda \mathbf{M} - \mathbf{A}$ singular \Rightarrow *autovalor generalizado* é solução da equação polimomial $\det(\lambda \mathbf{M} - \mathbf{A}) = 0$.

Suponha \mathbf{M} p.d. Então, $\exists \mathbf{R}$ inversível t.q. $\mathbf{M} = \mathbf{R}^T \mathbf{R}$. Mudança de variáveis: $\mathbf{y} = \mathbf{R}\mathbf{x}$ implica em:

$$\mathbf{Ax} = \lambda \mathbf{Mx} = \lambda \mathbf{R}^T \mathbf{R}\mathbf{x} \rightarrow \mathbf{AR}^{-1} \mathbf{y} = \lambda \mathbf{R}^T \mathbf{y}.$$

Seja $\mathbf{C} := \mathbf{R}^{-1}$, pre-multiplicando por $(\mathbf{R}^T)^{-1} = \mathbf{C}^T$:

$$\mathbf{C}^T \mathbf{AC} = \lambda \mathbf{y}.$$

Os AV para este problema padrão (*com uma matriz*) coincidem com os AV do problema original $\mathbf{Ax} = \lambda \mathbf{Mx}$, e os autovetores são relacionados por $\mathbf{y}_j = \mathbf{R}\mathbf{x}_j$.

Propriedades do problema de autovalor generalizado

Pelas propriedades da matriz *simétrica* $\mathbf{C}^T \mathbf{A} \mathbf{C}$, temos as propriedades correspondentes de $\mathbf{A} \mathbf{x} = \lambda \mathbf{M} \mathbf{x}$:

1. Os autovalores generalizados são reais.
2. Os AV generalizados possuem os mesmos sinais que os AV (usuais) de \mathbf{A} (pela lei de inércia).
3. Os $\overrightarrow{\mathbf{AV}}$ de $\mathbf{C}^T \mathbf{A} \mathbf{C}$ podem ser escolhidos o.n., portanto os $\overrightarrow{\mathbf{AV}} \mathbf{x}_j$ satisfazem 'M-ortonormalidade':

$$\mathbf{x}_i^T \mathbf{M} \mathbf{x}_j = \mathbf{x}_i^T \mathbf{R}^T \mathbf{R} \mathbf{x}_j = \mathbf{y}_i^T \mathbf{y}_j = 1, \text{ se } i = j; 0, \text{ se } i \neq j.$$

4. Da mesma forma $\mathbf{x}_i^T \mathbf{A} \mathbf{x}_j = \lambda_j \mathbf{x}_i^T \mathbf{M} \mathbf{x}_j$: é igual a λ_j ou 0. Isto é, as matrizes \mathbf{A} e \mathbf{M} estão sendo **simultaneamente diagonalizados**, através de uma **congruência**. Se colocarmos os \mathbf{x}_i nas colunas de uma matriz \mathbf{X} , então $\mathbf{X}^T \mathbf{A} \mathbf{X} = \Lambda$ e $\mathbf{X}^T \mathbf{M} \mathbf{X} = \mathbf{I}$.

Conclusão final: Sempre que \mathbf{M} for p.d., o problema $\mathbf{A} \mathbf{x} = \lambda \mathbf{M} \mathbf{x}$ (do AV generalizado) possui comportamento parecido com o do problema usual $\mathbf{A} \mathbf{x} = \lambda \mathbf{x}$.

Princípio de mínimo para $\mathbf{Ax} = \mathbf{b}$

Energia = forma quadrática p.d., derivada é linear!

Se a matriz \mathbf{A} for simétrica p.d., então $P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{x}^T \mathbf{b}$ atinge seu mínimo no ponto onde $\mathbf{Ax} = \mathbf{b}$. Neste ponto, o valor mínimo atingido é $P_{\min} = -\frac{1}{2}\mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}$.

Aplicação desta observação: métodos de mínimo para resolver $\mathbf{Ax} = \mathbf{b}$ (gradiente conjugado, etc.), porque surgem \mathbf{A} s em aplicações que são simétricas, p.d., de dimensão elevada, porém esparsas (= poucos elementos não-nulos).

Princípio de mínimo para $\mathbf{Ax} = \lambda\mathbf{x}$

Função a ser minimizada não pode ser quadrática: levaria a um problema linear!

Teorema (ou princípio) de Rayleigh: Dada uma matriz $\mathbf{A} = \mathbf{A}^T$, o *quociente de Rayleigh* de \mathbf{A} ,

$$R_{\mathbf{A}}(\mathbf{x}) := \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}},$$

é minimizado (resp. maximizado) pelo $\overrightarrow{\mathbf{AV}}$ \mathbf{x}_1 (resp. \mathbf{x}_n) correspondente ao menor AV λ_1 (resp. maior λ_n) e seu valor mínimo (resp. máximo) é dado por λ_1 (resp. λ_n). Ou seja:

$$\forall \mathbf{x}, \lambda_{\min}(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{x}) \leq \lambda_{\max}(\mathbf{A}),$$

e $R_{\mathbf{A}}(\mathbf{x}_1) = \lambda_{\min}(\mathbf{A})$, $R_{\mathbf{A}}(\mathbf{x}_n) = \lambda_{\max}(\mathbf{A})$, onde os autovalores de \mathbf{A} são colocados em ordem crescente $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, e os \mathbf{x}_i são os autovetores correspondentes.