Newton Method on Brockett Function

Wen Perng

03/07/2024

Let $\mathbf{A}$, $\mathbf{B} \in \mathbb{R}^{m\times m}$ be symmetric positive definite. We define the Brockett function as

$$f : \mathrm{O}(m) \to \mathbb{R} \qquad \mathbf{X} \mapsto \frac{1}{2}\mathrm{Tr}\left\{\mathbf{X}^\top \mathbf{A}\mathbf{X}\mathbf{B}\right\}. \tag{1}$$

Further, a local parameterisation of $\mathrm{O}(m)$ at $\mathbf{X}$ is given as

$$\zeta_{\mathbf{X}} : \mathrm{Skew}(m) \to \mathrm{O}(m) \qquad \mathbf{\Omega} \mapsto \mathbf{X}\cdot \exp(\mathbf{\Omega}). \tag{2}$$

Question 2:
Develop a Newton-like method for minimising f.

Note that

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_0 \zeta_{\mathbf{X}}(t\mathbf{\Omega}) = \mathbf{X}\mathbf{\Omega}, \quad \left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_0 \zeta_{\mathbf{X}}(t\mathbf{\Omega}) = \mathbf{X}\mathbf{\Omega}^2.$$

First we compute the gradient to be

$$\begin{aligned} \mathrm{D}\left(f\circ\zeta_{\mathbf{X}}(0)\right)\mathbf{\Omega} &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_0 \frac{1}{2}\mathrm{Tr}\left\{\zeta_{\mathbf{X}}^\top A\zeta_{\mathbf{X}} B\right\} = \mathrm{Tr}\left\{\mathbf{\Omega}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B}\right\} \\ \Rightarrow \mathrm{grad} f(\mathbf{X}\circ\zeta_{\mathbf{X}})(0) &= \mathrm{skew}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B}). \end{aligned}$$

Next, we find the Hessian by calculating the second derivative:

$$\mathrm{D}^2\left(f\circ\zeta_{\mathbf{X}}(0)\right)(\mathbf{\Omega},\mathbf{\Omega}) = \left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_0 \frac{1}{2}\mathrm{Tr}\left\{\zeta_{\mathbf{X}}^\top A\zeta_{\mathbf{X}} B\right\} = \mathrm{Tr}\left\{{\mathbf{\Omega}^\top}^2\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{\Omega}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{B}\right\}.$$

Applying the polarization identity to obtain the bilinear form on $\mathbf{\Omega}$, $\mathbf{\Lambda} \in \mathrm{Skew}(m)$:

$$\begin{aligned} \mathrm{D}^2\left(f\circ\zeta_{\mathbf{X}}(0)\right)(\mathbf{\Omega},\mathbf{\Lambda}) &= \frac{1}{2}\mathrm{Tr}\Big\{\underbrace{\mathbf{\Omega}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Lambda}\mathbf{B}}_{=\mathbf{B}\mathbf{\Lambda}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}} + \mathbf{\Lambda}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{B} + \mathbf{\Omega}^\top\mathbf{\Lambda}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B} + \underbrace{\mathbf{\Lambda}^\top\mathbf{\Omega}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B}}_{=\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{\Lambda}}\Big\} \\ &= \frac{1}{2}\mathrm{Tr}\left\{2\mathbf{\Lambda}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{B} - \mathbf{\Lambda}^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{\Omega} - \mathbf{\Lambda}^\top\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\right\} \\ &= \mathrm{Tr}\left\{\mathbf{\Lambda}^\top\left(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{B} - \mathrm{sym}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B})\mathbf{\Omega}\right)\right\} \\ \Rightarrow \mathrm{hess}(f\circ\zeta_{\mathbf{X}})(0)\mathbf{\Omega} &= \mathrm{skew}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{B} - \mathrm{sym}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B})\mathbf{\Omega}). \end{aligned}$$

For us to find the Newton's direction, we need to solve the Newton's equation of

$$\begin{aligned}\operatorname{hess}(f\circ\zeta_{\mathbf{X}})(0)\mathbf{\Omega} &= -\operatorname{grad} f(\mathbf{X}\circ\zeta_{\mathbf{X}})(0)\\ \operatorname{skew}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{\Omega}\mathbf{B}-\operatorname{sym}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B})\mathbf{\Omega}) &= -\operatorname{skew}(\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{B}).\end{aligned}$$

Let us define $\mathcal{A} := \mathbf{X}^\top\mathbf{A}\mathbf{X}$, and $\mathbf{Z} := -2\operatorname{skew}(\mathcal{A}B)$, then we have

$$\begin{aligned}\operatorname{skew}(\mathcal{A}\mathbf{\Omega}\mathbf{B}-\operatorname{sym}(\mathcal{A}B)\mathbf{\Omega}) &= \mathbf{Z}\\ \mathcal{A}\mathbf{\Omega}\mathbf{B}+\mathbf{B}\mathbf{\Omega}\mathcal{A}-\operatorname{sym}(\mathcal{A}\mathbf{B})\mathbf{\Omega}-\mathbf{\Omega}\operatorname{sym}(\mathcal{A}\mathbf{B}) &= \mathbf{Z}.\end{aligned} \tag{3}$$

How do we solve for Ω? One might immediately shout *vectorization*! The matrix equation quickly turns into a matrix-vector equation:

$$\underbrace{\left(\mathcal{A}\otimes\mathbf{B}+\mathbf{B}\otimes\mathcal{A}-\mathbb{1}\otimes\operatorname{sym}(\mathcal{A}\mathbf{B})-\operatorname{sym}(\mathcal{A}\mathbf{B})\otimes\mathbb{1}\right)}_{=:\mathcal{M}}\vec{\mathbf{\Omega}}=\vec{\mathbf{Z}}, \tag{4}$$

where $\vec{\mathbf{X}} := \operatorname{vec}(\mathbf{X})$, and $\operatorname{vec}(\mathbf{AXB}) = (\mathbf{B}^\top\otimes\mathbf{A})\vec{\mathbf{X}}$. Sadly, the matrix $\mathcal{M}$ is not invertible:

Lemma. ***$\mathcal{M}$ has 0 as an eigenvalue. Hence it is non-invertible.***
Proof. Denote the linear operator $\mathcal{M}^\downarrow : \mathbb{R}^{m\times m}\to\mathbb{R}^{m\times m}$ as

$$\mathcal{M}^\downarrow(\mathbf{\Omega}) = \mathcal{A}\mathbf{\Omega}\mathbf{B}+\mathbf{B}\mathbf{\Omega}\mathcal{A}-\operatorname{sym}(\mathcal{A}\mathbf{B})\mathbf{\Omega}-\mathbf{\Omega}\operatorname{sym}(\mathcal{A}\mathbf{B}),$$

its operator lift is $\mathcal{M}$. Note that it has the null space of $\mathbb{1}$:

$$\begin{aligned}\mathcal{M}^\downarrow(\mathbb{1}) = \mathcal{A}\mathbf{B}+\mathbf{B}\mathcal{A}-2\cdot\operatorname{sym}(\mathcal{A}\mathbf{B}) &= 0\\ \mathcal{M}\vec{\mathbb{1}} = \operatorname{vec}(\mathcal{M}^\downarrow(\mathbb{1})) &= 0.\end{aligned}$$

Hence, $\mathcal{M}$ has 0 as an eigenvalue. ■

We need another method to solve for the matrix equation $\mathcal{M}\vec{\mathbf{\Omega}}=\vec{\mathbf{Z}}$.

Here I shall present three methods to solve the Newton's equation above.

Method 1.

Note that we have yet utilized the property that $\mathbf{\Omega}$ is skew-symmetric. Let us introduce the commutation matrix $\mathbf{K}$ such that it satisfies

$$\mathbf{K}\cdot\operatorname{vec}(\mathbf{X}) = \operatorname{vec}(\mathbf{X}^\top), \tag{5}$$

all skew-symmetric matrices satisfy

$$\mathbf{K}\cdot\operatorname{vec}(\mathbf{\Omega}) = \operatorname{vec}(\mathbf{\Omega}^\top) = -\operatorname{vec}(\mathbf{\Omega}). \tag{6}$$

Let us spectrally decompose $\mathcal{M}$ into its eigenvalues λ_i and the corresponding mutually orthonormal (wrt Euclidean metric) eigenvectors $\vec{\mathbf{V}}_i$. Since $\mathcal{M}$ is symmetric, such decomposition into an orthonormal set of eigenvectors is possible. Out of all the m^2 eigenvectors, there are exactly $\frac{m(m-1)}{2}$ of them forming a complete basis for the vectorization of skew-symmetric matrices, and is identified using $\mathbf{K}$.

Lemma. ***The matrix $\mathcal{M}$ has eigenvectors spanning the vectorization of skew-symmetric matrices, and these eigenvectors have non-zero eigenvalues.***
Proof. I have no idea..., MATLAB says I'm correct though. Note that $\mathbf{A}$ and $\mathbf{B}$ are required to have distinct eigenvalues respectively for the lemma to hold (sufficiently). Further,

$$\mathcal{M}^\downarrow\left(\operatorname{Skew}(m)\right)\subseteq\operatorname{Skew}(m).$$

Hence we choose $\{\vec{\mathbf{W}}_i\} \subset \{\vec{\mathbf{V}}_j\}$ such that $\mathbf{K}\vec{\mathbf{W}}_i = -\vec{\mathbf{W}}_i$, $\langle \vec{\mathbf{W}}_i, \vec{\mathbf{W}}_i \rangle = 1$, i.e. $\vec{\mathbf{V}}_i = \vec{\mathbf{W}}_i$ $(\forall i = 1, \cdots, \frac{m(m-1)}{2})$, and they form a complete basis of skew-symmetric matrices. Thus the Newton's equation becomes

$$\mathcal{M}\vec{\mathbf{\Omega}} = \left(\sum_j \lambda_j \vec{\mathbf{V}}_j \vec{\mathbf{V}}_j^\top\right)\left(\sum_i \omega_i \vec{\mathbf{W}}_i\right) = \sum_i \langle \vec{\mathbf{Z}}, \vec{\mathbf{W}}_i \rangle \vec{\mathbf{W}}_i = \vec{\mathbf{Z}}.$$

It is then easy to check that the solution will be

$$\vec{\mathbf{\Omega}} = \sum_i \frac{\langle \vec{\mathbf{Z}}, \vec{\mathbf{W}}_i \rangle}{\lambda_i} \vec{\mathbf{W}}_i. \tag{7}$$

Lastly, by unvectorize $\mathbf{\Omega} = \mathrm{vec}^{-1}(\vec{\mathbf{\Omega}})$, we obtain the solution. The Newton-like method can be readily applied. ■

Method 2.

Since $\mathbf{\Omega}$ is skew-symmetric, some of the equations in $\mathcal{M}$ is redundant, and we should remove them to obtain a smaller matrix $\mathcal{N}$ that is invertible.

Which of the elements in $\mathcal{M}$ is redundant? Let us take a look at low-dimensional cases for intuition:

(a) $m = 2$,

$$\mathcal{M}\vec{\mathbf{\Omega}} = \begin{bmatrix} * & * & * & * \\ * & a & c & * \\ * & b & d & * \\ * & * & * & * \end{bmatrix}\begin{bmatrix} 0 \\ w \\ -w \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ z \\ -z \\ 0 \end{bmatrix} = \vec{\mathbf{Z}}$$

(b) $m = 3$, $(1) \mapsto (1) - (4)$, $(2) \mapsto (2) - (5)$, $(3) \mapsto (3) - (6)$

$$\mathcal{M}\vec{\mathbf{\Omega}} = \begin{bmatrix} * & * & * & * & * & * & * & * & * \\ * & | & | & | & * & | & | & | & * \\ * & (1) & (2) & (4) & * & (3) & (5) & (6) & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ * & | & | & | & * & | & | & | & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \end{bmatrix}\begin{bmatrix} 0 \\ a \\ b \\ -a \\ 0 \\ c \\ -b \\ -c \\ 0 \end{bmatrix} = \vec{\mathbf{Z}}$$

By combining/removing redundant equations in $\mathcal{M}$ into $\frac{m(m-1)}{2}$ independent ones, the remaining matrix $\mathcal{N}$ is invertible. Hence the Newton's direction is obtained. ■

Note that we have not proven the invertibility of the reduced matrix $\mathcal{N}$.

Method 3. The vector $\vec{\mathbf{\Omega}}$ satisfies two constraints:

$$\begin{cases} \mathcal{M}\vec{\mathbf{\Omega}} = \vec{\mathbf{Z}} \\ (\mathbb{1} + \mathbf{K})\vec{\mathbf{\Omega}} = 0 \end{cases}, \tag{8}$$

where the second constraint uses the skew-symmetric property of $\mathbf{\Omega}$. We can combine the two equations into a single one as:

$$\tilde{\mathcal{M}}\vec{\mathbf{\Omega}} := \begin{bmatrix} \mathcal{M} \\ \mathbb{1} + \mathbf{K} \end{bmatrix} \vec{\mathbf{\Omega}} = \begin{bmatrix} \vec{\mathbf{Z}} \\ 0 \end{bmatrix} =: \tilde{\vec{\mathbf{Z}}}, \tag{9}$$

this can be solved by applying the pseudo-inverse to $\tilde{\mathcal{M}}$, denoted by $\tilde{\mathcal{M}}^{+}$, obtaining the solution of

$$\vec{\mathbf{\Omega}} = \tilde{\mathcal{M}}^{+}\tilde{\vec{\mathbf{Z}}}. \tag{10}$$

The Newton's method can be applied readily. ■

Note that by using the pseudo-inverse to solve Equation 9, the solution we obtained in Equation 10 are the min-square-error solution. The fact that the two results coincide is not yet proven.

Lemma. ***The solution to Equation 9 and Equation 10 coincides.***
Proof. IDK

Question 4:
Let $\mathbf{B} \in \mathbb{R}^{m\times m}$ be further assumed to be diagonal. Develop a simpler approximate Newton-like method for minimising f.

For the case of $\mathbf{B}$ being diagonal, the method for obtaining the Newton's direction for the approximate Newton-like method is even computationally simpler. At the critical point $\mathbf{X}^*$,

$$\operatorname{grad} f(\mathbf{X}^*) = 0 = \operatorname{skew}(\mathbf{X^*}^\top \mathbf{A}\mathbf{X}^*\mathbf{B}) \Leftrightarrow \mathbf{X^*}^\top \mathbf{A}\mathbf{X}^* \text{ and } \mathbf{B} \text{ commutes} \Leftrightarrow \mathcal{A}^* := \mathbf{X^*}^\top \mathbf{A}\mathbf{X}^* \text{ is diagonal.}$$

For the approximate Newton-like method, we want to find a matrix $\mathbf{H}(\mathbf{X})$ satisfying

$$\begin{aligned}
\mathbf{H}(\mathbf{X}^*)\mathbf{\Omega} \equiv \operatorname{hess}(f\circ\zeta_{\mathbf{X}^*})(0)\mathbf{\Omega} &= \operatorname{skew}(\mathcal{A}^*\mathbf{\Omega}\mathbf{B} - \operatorname{sym}(\mathcal{A}^*\mathbf{B})\mathbf{\Omega}) \\
&= \operatorname{skew}(\mathcal{A}^*\mathbf{\Omega}\mathbf{B} - \mathcal{A}^*\mathbf{B}\mathbf{\Omega}) \\
&= \operatorname{skew}\left(\left[\mathcal{A}^*_{ii}\mathbf{\Omega}_{ij}\mathbf{B}_{jj} - \mathcal{A}^*_{ii}\mathbf{B}_{ii}\mathbf{\Omega}_{ij}\right]\right) = \operatorname{skew}\left(\left[\mathcal{A}^*_{ii}(\mathbf{B}_{jj} - \mathbf{B}_{ii})\mathbf{\Omega}_{ij}\right]\right) \\
&= \frac{1}{2}\left[\mathcal{A}^*_{ii}(\mathbf{B}_{jj}-\mathbf{B}_{ii})\mathbf{\Omega}_{ij} - \mathcal{A}^*_{jj}(\mathbf{B}_{ii}-\mathbf{B}_{jj})\mathbf{\Omega}_{ji}\right] \\
&= \left[-\frac{1}{2}(\mathcal{A}^*_{ii}-\mathcal{A}^*_{jj})(\mathbf{B}_{ii}-\mathbf{B}_{jj})\mathbf{\Omega}_{ij}\right].
\end{aligned}$$

Henceforth, I have chosen

$$\mathbf{H}(\mathbf{X})\mathbf{\Omega} = \left[-\frac{1}{2}(\mathcal{A}_{ii}-\mathcal{A}_{jj})(\mathbf{B}_{ii}-\mathbf{B}_{jj})\mathbf{\Omega}_{ij}\right], \tag{11}$$

where $\mathcal{A} = \mathbf{X}^\top\mathbf{A}\mathbf{X}$. Solving the Newton's equation also becomes extremely simple:

$$\begin{aligned}
\left[-\operatorname{hess}(f\circ\zeta_{\mathbf{X}})(0)\mathbf{\Omega}\right]_{ij} = \frac{1}{2}(\mathcal{A}_{ii}-\mathcal{A}_{jj})(\mathbf{B}_{ii}-\mathbf{B}_{jj})\mathbf{\Omega}_{ij} &= \left[\operatorname{skew}(\mathcal{A}\mathbf{B})\right]_{ij} = \left[\operatorname{grad}(f\circ\zeta_{\mathbf{X}})(0)\right]_{ij} \\
\Rightarrow \mathbf{\Omega}_{ij} &= \begin{cases} \dfrac{2\left[\operatorname{skew}(\mathcal{A}\mathbf{B})\right]_{ij}}{\left([\mathbf{X}^\top\mathbf{A}\mathbf{X}]_{ii} - [\mathbf{X}^\top\mathbf{A}\mathbf{X}]_{jj}\right)(\mathbf{B}_{ii}-\mathbf{B}_{jj})} & \text{, for } i \neq j; \\ 0 & \text{, for } i = j. \end{cases} \quad \blacksquare
\end{aligned} \tag{12}$$

Note that for obtaining the right solutions, it is required that the eigenvalues of $\mathbf{B}$ are all distinct. This is, I *assume*, implicitly implied when we want our Hessian to be non-singular, also implying that all the critical points are isolated.

Question 3:
Develop an approximate Newton-like method for minimising f.

At the critical point $\mathbf{X}^*$, it remains that $\mathcal{A}^* := \mathbf{X}^{*\top}\mathbf{A}\mathbf{X}^*$ and $\mathbf{B}$ commutes. For the approximate Newton-like method, we want to find a matrix $\mathbf{H}(\mathbf{X})$ satisfying

$$\begin{aligned}\mathbf{H}(\mathbf{X}^*)\boldsymbol{\Omega} \equiv \mathrm{hess}(f \circ \zeta_{\mathbf{X}^*})(0)\boldsymbol{\Omega} &= \mathrm{skew}(\mathcal{A}^*\boldsymbol{\Omega}\mathbf{B} - \mathrm{sym}(\mathcal{A}^*\mathbf{B})\boldsymbol{\Omega}) \\ &= \mathrm{skew}(\mathcal{A}^*\boldsymbol{\Omega}\mathbf{B} - \mathcal{A}^*\mathbf{B}\boldsymbol{\Omega})\end{aligned}$$

Henceforth, I have chosen

$$\mathbf{H}(\mathbf{X})\boldsymbol{\Omega} = \mathrm{skew}(\mathcal{A}\boldsymbol{\Omega}\mathbf{B} - \mathcal{A}\mathbf{B}\boldsymbol{\Omega}) \tag{13}$$

where $\mathcal{A} = \mathbf{X}^\top\mathbf{A}\mathbf{X}$. Procedure of solving the Newton's equation is the same as in question 2.