Close form formula of $\|\mathbf{I} + k\mathbf{u}\mathbf{v}^\top\|_2$

$$= \max\left\{1, \sqrt{1 + k\langle \mathbf{u}, \mathbf{v}\rangle + \frac{k^2}{2} + \frac{k}{2}\sqrt{k^2 + 4 + 4k\langle \mathbf{u}, \mathbf{v}\rangle}}\right\}$$

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : Janurary 11, 2019
Last update : January 11, 2019

Setting

Given two unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, find the operator norm of the matrix

$$\mathbf{X} := \mathbf{I} + k\mathbf{u}\mathbf{v}^\top,$$

where $\mathbf{X}$ is the sum of identity plus a rank-1 perturbation :

- $\mathbf{u}\mathbf{v}^\top$ is the outer product of the vector pair, which is a rank-1 perturbation matrix
- k control the "size" of the rank-1 perturbation matrix : it control how far $\mathbf{X}$ is drifted away from $\mathbf{I}$ due to the perturbation

Note : we can always let $k \geq 0$ since we can always absorb the negative sign into $\mathbf{u}$ or $\mathbf{v}$.

Now we want to know the behaviour of $\mathbf{X}$ being drifted away from $\mathbf{I}$ by the perturbation. That is, what is the relationship between the norm of $\mathbf{X}$ (denoted as $\|\mathbf{X}\|$) and the given k, $\mathbf{u}$, $\mathbf{v}$?

In other words, we want to characterize the norm of a rank-1 perturbation.

Characterization of norm of a rank-1 perturbation

We can completely characterize the behaviour of $\mathbf{X}$ being drifted away by the perturbation as follows.

Given $\mathbf{u}, \mathbf{v}$, we know the angle between $\mathbf{u}$ and $\mathbf{v}$: $h = \cos\theta = \langle \mathbf{u}, \mathbf{v} \rangle$.

In the case of operator norm, we have the following :

$$\|\mathbf{X}\| = \sup_{\|\mathbf{y}\|=1} \|\mathbf{X}\mathbf{y}\| = \max\left\{1, \left(1 + kh + \frac{k^2}{2} + \frac{k}{2}\sqrt{4 + 4kh + h^2}\right)^{\frac{1}{2}}\right\}.$$

This document : prove this formula.

Key idea of the proof : orthonormal matrix . . . 1/3

The key idea as well as the starting point of the proof is the unitary invariant property of operator norm : for all orthonormal matrix $\mathbf{Q}$:

$$\|\mathbf{Q}^\top \mathbf{X} \mathbf{Q}\| = \|\mathbf{X}\|.$$

Put $\mathbf{X} = \mathbf{I} + k\mathbf{u}\mathbf{v}^\top$, we have

$$\begin{aligned} \|\mathbf{X}\| = \|\mathbf{Q}^\top \mathbf{X} \mathbf{Q}\| &= \|\mathbf{Q}^\top (\mathbf{I} + k\mathbf{u}\mathbf{v}^\top)\mathbf{Q}\| \\ &\overset{\mathbf{Q}^\top\mathbf{Q}=\mathbf{I}}{=} \|\mathbf{I} + k\mathbf{Q}^\top \mathbf{u}\mathbf{v}^\top \mathbf{Q}\| \\ &= \|\mathbf{I} + k\mathbf{Q}^\top \mathbf{u}(\mathbf{Q}^\top \mathbf{v})^\top\| \\ &= \|\mathbf{I} + k\mathbf{u}'\mathbf{v}'^\top\|, \end{aligned}$$

where $\mathbf{u}' = \mathbf{Q}^\top \mathbf{u}$ and $\mathbf{v}' = \mathbf{Q}^\top \mathbf{v}$ are the rotated vectors $\mathbf{u}, \mathbf{v}$. The rotation is $\mathbf{Q}$. The expression above tells that such rotation does not change the norm of $\mathbf{X}$.

As $\mathbf{Q}$ is any arbitrary, we have the freedom to design a special $\mathbf{Q}$ to fit out purpose : we can just let $\mathbf{Q}$ be a orthonormal matrix with the first column being as $\mathbf{u}$ (or $\mathbf{v}$).

Key idea of the proof : orthonormal matrix . . . 2/3

Suppose $\mathbf{Q}$ has the structure

$$\mathbf{Q} = \begin{bmatrix} | & | & | & \dots & | \\ \mathbf{u} & \mathbf{q}_2 & \mathbf{q}_3 & \dots & \mathbf{q}_n \\ | & | & | & \dots & | \end{bmatrix},$$

where $\mathbf{q}_2, \mathbf{q}_3, \dots, \mathbf{q}_n$ are $n-1$ unit vectors in $\mathbb{R}^n$ that are orthogonal to $\mathbf{u}$. Then

$$\mathbf{Q}^\top \mathbf{u} = \begin{bmatrix} - & \mathbf{u} & - \\ - & \mathbf{q}_2 & - \\ - & \mathbf{q}_3 & - \\ \vdots & \vdots & \vdots \\ - & \mathbf{q}_n & - \end{bmatrix} \mathbf{u} = \begin{bmatrix} \langle \mathbf{u}, \mathbf{u} \rangle \\ \langle \mathbf{q}_2, \mathbf{u} \rangle \\ \langle \mathbf{q}_3, \mathbf{u} \rangle \\ \vdots \\ \langle \mathbf{q}_n, \mathbf{u} \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_1, \tag{1}$$

where $\mathbf{e}_i$ stand for the i^{th} standard basis vector.

Geometrically, the expression $\mathbf{Q}^\top \mathbf{u} = \mathbf{e}_1$ means $\mathbf{Q}$ rotates the unit vector $\mathbf{u}$ to give the vector along x-axis.

Now consider $\mathbf{Q}^\top \mathbf{y}$. For $\mathbf{Q} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{u} & \mathbf{q}_2 & \dots & \mathbf{q}_n \\ | & | & \dots & | \end{bmatrix}$, we have

$$\mathbf{Q}^\top \mathbf{v} = \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{q}_2, \mathbf{v} \rangle \\ \vdots \\ \langle \mathbf{q}_n, \mathbf{v} \rangle \end{bmatrix} = \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle \\ \bar{\mathbf{v}} \end{bmatrix}, \text{ where } \bar{\mathbf{v}} = \begin{bmatrix} \langle \mathbf{q}_2, \mathbf{v} \rangle \\ \vdots \\ \langle \mathbf{q}_n, \mathbf{v} \rangle \end{bmatrix} \in \mathbb{R}^{n-1}.$$

We have

$$\|\mathbf{Q}^\top \mathbf{v}\|^2 = \langle \mathbf{u}, \mathbf{v} \rangle^2 + \|\bar{\mathbf{v}}\|^2.$$

By the fact that $\mathbf{v}$ has unit norm and norm is unitary invariant : we have $\|\mathbf{v}\| = \|\mathbf{Q}^\top \mathbf{v}\| = 1$ and also

$$\|\bar{\mathbf{v}}\|^2 = 1 - \langle \mathbf{u}, \mathbf{v} \rangle^2. \tag{2}$$

Now we can begin the proof.

The proof ... 1/4

Let $\mathbf{Q} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{u} & \mathbf{q}_2 & \dots & \mathbf{q}_n \\ | & | & \dots & | \end{bmatrix}$, where $\mathbf{q}_2, \dots, \mathbf{q}_n \in \mathbb{R}^n$ are $n-1$ unit vectors orthogonal to $\mathbf{u}$. Then

$$\|\mathbf{X}\| = \|\mathbf{I} + k\mathbf{Q}^\top \mathbf{u}(\mathbf{Q}^\top \mathbf{v})^\top\| \overset{(1)}{=} \|\mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top\|.$$

Recall $\|\mathbf{X}\|^2 = \sigma^2_{\max}(\mathbf{X}) = \lambda_{\max}(\mathbf{X}\mathbf{X}^\top)$, consider $\mathbf{X}\mathbf{X}^\top$

$$\begin{aligned}
\mathbf{X}\mathbf{X}^\top &= (\mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top)(\mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top)^\top \\
&= (\mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top)(\mathbf{I} + k\mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top) \\
&= \mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top + k\mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top k\mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top \\
&= \mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top + k\mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top + k^2\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top \mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top.
\end{aligned}$$

As $(\mathbf{Q}^\top \mathbf{v})^\top \mathbf{Q}^\top \mathbf{v} = \mathbf{v}^\top \underbrace{\mathbf{Q}\mathbf{Q}^\top}_{\mathbf{I}} \mathbf{v} = \mathbf{v}^\top \mathbf{v} = 1$, then

$$\|\mathbf{X}\|^2 = \lambda_{\max}\left(\mathbf{X}\mathbf{X}^\top\right) = \lambda_{\max}\left(\mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top + k\mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top + k^2\mathbf{e}_1\mathbf{e}_1^\top\right).$$

The proof . . . 2/4

Let $\mathbf{I}_{n-1}$, $\mathbf{O}_{n-1}$ be the identity and zero matrix of order $n-1$ respectively.

Let $\mathbf{0}$ be zero vector in $\mathbb{R}^{n-1}$ so $\mathbf{e}_1 = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}$. With $\mathbf{Q}^\top \mathbf{v} = \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle \\ \bar{\mathbf{v}} \end{bmatrix}$:

$$
\begin{aligned}
& \mathbf{X}\mathbf{X}^\top \\
= \;& \mathbf{I} + k\mathbf{e}_1(\mathbf{Q}^\top \mathbf{v})^\top + k\mathbf{Q}^\top \mathbf{v}\mathbf{e}_1^\top + k^2 \mathbf{e}_1 \mathbf{e}_1^\top \\
= \;& \mathbf{I} + k \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle \\ \bar{\mathbf{v}} \end{bmatrix}^\top + k \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle \\ \bar{\mathbf{v}} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}^\top + k^2 \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}^\top \\
= \;& \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_{n-1} \end{bmatrix} + k \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle & \bar{\mathbf{v}}^\top \\ \mathbf{0} & \mathbf{O}_{n-1} \end{bmatrix} + k \begin{bmatrix} \langle \mathbf{u}, \mathbf{v} \rangle & \mathbf{0}^\top \\ \bar{\mathbf{v}} & \mathbf{O}_{n-1} \end{bmatrix} + k^2 \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{O}_{n-1} \end{bmatrix} \\
= \;& \begin{bmatrix} 1 + 2k\langle \mathbf{u}, \mathbf{v} \rangle + k^2 & k\bar{\mathbf{v}}^\top \\ k\bar{\mathbf{v}} & \mathbf{I}_{n-1} \end{bmatrix}.
\end{aligned}
$$

So we expressed $\|\mathbf{X}\|^2$ as the largest eigenvalue of the 2-by-2 block matrix

$$
\|\mathbf{X}\|^2 \;=\; \lambda_{\max}\left(\begin{bmatrix} 1 + 2k\langle \mathbf{u}, \mathbf{v} \rangle + k^2 & k\bar{\mathbf{v}}^\top \\ k\bar{\mathbf{v}} & \mathbf{I}_{n-1} \end{bmatrix} \right).
$$

With $\|\mathbf{X}\|^2 = \lambda_{\max}\left(\begin{bmatrix} 1 + 2k\langle\mathbf{u},\mathbf{v}\rangle + k^2 & k\bar{\mathbf{v}}^\top \\ k\bar{\mathbf{v}} & \mathbf{I}_{n-1}\end{bmatrix}\right)$, we compute the eigenvalue by solving the characteristic equation

$$0 = \det\left(\begin{bmatrix} \lambda - (1 + 2k\langle\mathbf{u},\mathbf{v}\rangle + k^2) & -k\bar{\mathbf{v}}^\top \\ -k\bar{\mathbf{v}} & (\lambda - 1)\mathbf{I}_{n-1}\end{bmatrix}\right). \tag{3}$$

Using the formula of determinant of block matrix

$$\det\left(\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix}\right) = \det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})\det(\mathbf{D}),\ \ \mathbf{D} \text{ is non-singular.}$$

(3) becomes

$$\begin{aligned} 0 &= \det\left(\lambda - (1 + 2k\langle\mathbf{u},\mathbf{v}\rangle + k^2) - \frac{k^2}{\lambda - 1}\underbrace{\bar{\mathbf{v}}^\top\bar{\mathbf{v}}}_{\|\bar{\mathbf{v}}\|^2}\right)\det\left((\lambda - 1)\mathbf{I}_{n-1}\right) \\ &\overset{(2)}{=} \det\left(\lambda - (1 + 2k\langle\mathbf{u},\mathbf{v}\rangle + k^2) - \frac{k^2}{\lambda - 1}(1 - \langle\mathbf{u},\mathbf{v}\rangle^2)\right)(\lambda - 1) \\ &= \left(\lambda - (1 + 2k\langle\mathbf{u},\mathbf{v}\rangle + k^2) - \frac{k^2}{\lambda - 1}(1 - \langle\mathbf{u},\mathbf{v}\rangle^2)\right)(\lambda - 1). \end{aligned}$$

Let $h = \langle \mathbf{u}, \mathbf{v} \rangle$, solving the last equation means

$$\lambda - (1 + 2kh + k^2) - \frac{k^2}{\lambda - 1}(1 - h^2) = 0 \qquad \text{or} \qquad \lambda = 1$$

That is, the eigenvalues of $\mathbf{X}\mathbf{X}^\top$ are either 1 or λ^*, where λ^* is the root of the first equation above, which can be shown to be

$$\lambda^* = 1 + kh + \frac{k^2}{2} \pm \frac{k}{2}\sqrt{k^2 + 4 + 4kh}.$$

Note : a n-by-n matrix have n eigenvalue. The eigenvalue $\lambda = 1$ has the multiplicity of $n - 1$, due to the matrix $\mathbf{I}_{n-1}$. With these

$$\|\mathbf{X}\|^2 = \lambda_{\max}(\mathbf{X}\mathbf{X}^\top) = \max\left(1, 1 + kh + \frac{k^2}{2} + \frac{k}{2}\sqrt{k^2 + 4 + 4kh}\right).$$

By $\|\mathbf{X}\| = \sqrt{\|\mathbf{X}\|^2}$, we have

$$\|\mathbf{X}\| = \|\mathbf{I} + k\mathbf{u}\mathbf{v}^\top\| = \max\left\{1, \left(1 + kh + \frac{k^2}{2} + \frac{k}{2}\sqrt{k^2 + 4 + 4kh}\right)^{\frac{1}{2}}\right\}. \quad \square$$

The roots of $\lambda - (1 + 2kh + k^2) - \frac{k^2}{\lambda - 1}(1 - h^2) = 0$

The step-by-step proof to show the roots are

$$\lambda^* = 1 + kh + \frac{k^2}{2} \pm \frac{k}{2}\sqrt{k^2 + 4 + 4kh}.$$

Multiply $\lambda - 1$:

$$\lambda^2 - (1 + 2kh + k^2)\lambda - \lambda + (1 + 2kh + k^2) - k^2(1 - h^2) = 0.$$

Re-arrange

$$\lambda^2 - (2 + 2kh + k^2)\lambda + (1 + 2kh + k^2h^2) = 0.$$

Apply quadratic formula

$$\lambda = \frac{(2 + 2kh + k^2) \pm \sqrt{(2 + 2kh + k^2)^2 - 4(1 + 2kh + k^2h^2)}}{2}.$$

Simplify and expand

$$\lambda = 1 + kh + \frac{k^2}{2} \pm \frac{\sqrt{(4 + 4k^2h^2 + k^4 + 8kh + 4k^2 + 4k^3h) - (4 + 8kh + 4k^2h^2)}}{2}.$$

Simplify

$$\lambda = 1 + kh + \frac{k^2}{2} \pm \frac{k}{2}\sqrt{k^2 + 4 + 4kh} \quad \square$$

Last page - summary

Given $k \geq 0$, two unit vector $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the operator norm of the matrix $\mathbf{X} := \mathbf{I} + k\mathbf{u}\mathbf{v}^\top$ is

$$\|\mathbf{X}\| = \max\left\{1, \left(1 + kh + \frac{k^2}{2} + \frac{k}{2}\sqrt{k^2 + 4 + 4kh}\right)^{\frac{1}{2}}\right\}.$$

Special case : $\mathbf{u}, \mathbf{v}$ are orthogonal, $h = \langle \mathbf{u}, \mathbf{v} \rangle = 0$

$$\|\mathbf{X}\| = \max\left\{1, \sqrt{1 + \frac{k^2}{2} + \frac{k}{2}\sqrt{k^2 + 4}}\right\} \overset{k \geq 0}{=} \sqrt{1 + \frac{k^2}{2} + \frac{k}{2}\sqrt{k^2 + 4}}.$$

Note that the function $\sqrt{1 + x}$ is concave so we can upper bound it by it's first order Taylor approximation $1 + \dfrac{x}{2}$.

End of document