On bounding the Thompson metric by Schatten norms

David A. Snyder*

Abstract: The Thompson metric provides key geometric insights in the study of non-linear matrix equations and in many optimization problems. However, knowing that an approximate solution is within $d_T$ units, in the Thompson metric, of the actual solution provides little insight into how good the approximation is as a matrix or vector approximation. That is, bounding the Thompson metric between an approximate and accurate solution to a problem does not provide obvious bounds either for the spectral or the Frobenius norm, both Schatten norms, of the difference between the approximation and accurate solution. This paper reports such an upper bound, namely that $\|X - Y\|_F \leq 2^k \left( \frac{e^{d-1}}{e^{d-1}+1} \right) \max \{ \|X\|_F, \|Y\|_F \}$ where $\|\cdot\|_F$ denotes the Schatten $p$-norm and $d$ denotes the Thompson metric between $X$ and $Y$. Furthermore, a more geometric proof leads to a slightly better bound in the case of the Frobenius norm, $\|X - Y\|_F \leq \frac{(e^{d-1})}{\sqrt{e^{d+1}}} \sqrt{\|X\|_F^2 + \|Y\|_F^2} \leq 2^k \left( \frac{e^{d-1}}{e^{d-1}+1} \right) \max \{ \|X\|_F, \|Y\|_F \}$.

Subjects: Science; Mathematics & Statistics; Advanced Mathematics; Algebra; Linear & Multilinear Algebra; Analysis - Mathematics; Functional Analysis; Geometry

Keywords: Schatten Norm; Frobenius Norm; Thompson metric; Matrix equations

AMS Classification Codes: 15A60; 15B48; 15B57; 46B40

ABOUT THE AUTHOR

David A. Snyder is a professor in the Department of Chemistry at William Paterson University of NJ. He received a Bachelor of Science – double majoring in Mathematics and Biology – from the University of California, Irvine and a Ph.D. in Biochemistry from Rutgers University. Prior to starting his current position, Dr. Snyder was a post-doctoral researcher in Rafael Brüschweiler’s (then at Florida State University) research group. Dr. Snyder’s research interests include quantifying protein flexibility and covariance NMR, which comprises a family of techniques that use algebraic operations to enhance and combine NMR spectra. In the course of researching the accuracy of covariance NMR, Dr. Snyder found that many relevant algebraic results involved Thompson metrics rather than more intuitive metrics based on Schatten norms. The sparsity of mathematical literature relating Thompson metrics and Schatten norms prompted Dr. Snyder to prove the results presented in this paper.

PUBLIC INTEREST STATEMENT

Metrics are functions, mapping pairs of points to non-negative real numbers, which generalize the concept of distance to apply to abstract spaces. The Thompson metric provides critical geometric insights into dynamical systems, optimization problems and solving systems of equations. However, the Thompson metric is not an intuitive generalization of the concept of distance. The Thompson metric does provide an upper bound for more intuitive measurements of distance, such as those based on Schatten norms, but the currently known relation between Thompson metrics and more intuitive generalizations of distance is not always a tight bound. This paper presents a tighter bound relating metrics based on Schatten norms to Thompson metrics. The results in this paper can refine our geometrical understanding of problems arising in fluid mechanics, in geophysics as well as in robotics, and may improve assessments of the quality of data and image processing techniques.
1. Introduction

The Thompson metric is a variant of the Hilbert metric (Nussbaum & Walsh, 2004). The Hilbert metric generalizes the metric structure of hyperbolic geometry to the generalized concept of cones used in the study of Banach (complete normed vector) spaces, such as the space of Hermitian matrices. When applied to the unit disk, the Hilbert metric yields the Klein model of hyperbolic geometry, but when applied to a cone, such as the cone of positive definite or positive semidefinite matrices, the Hilbert metric is actually a pseudometric. A slight tweak of the Hilbert metric yields the Thompson (part) metric: the Thompson metric \( d_T(X, Y) \) is the minimal \( d_T = \log(\alpha) \) such that both \( aX - Y \) and \( aY - X \) are both positive semidefinite. The Thompson metric is well defined over the cone of positive definite matrices but may be infinite when applied to other matrices, such as positive semidefinite matrices.

The Thompson metric (Lemmens & Roelands, 2015; Nussbaum & Walsh, 2004) provides key geometric insights into the study of non-linear matrix equations. In particular, many flows, which in other metrics may not even be contractions, have well-characterized contraction rates in the Thompson metric (Lee & Lim, 2008). That flows arising in many non-linear optimization, filtering and control problems are contractions in the Thompson metric (Carli & Sepulchre, 2015; Del Moral, Kurtzmann, & Tugaut, 2017; Gaubert & Qu, 2014; Lawson & Lim, 2007; Qu, 2014) endows this metric with great utility. Applications of the Thompson metric range from proofs of the existence and uniqueness of positive definite solutions for many types of non-linear equations (Liao, et al., 2010) to non-linear optimization theory (Gaubert & Qu, 2014; Montrucchio, 1998) and nonlinear Perron-Frobenius theory (Lemmens & Nussbaum, 2012; Nussbaum, 1988). Relatedly, matrix bounds in the Löwner order characterize the error in approximate solutions to continuous algebraic Riccati equation (Zhang & Liu, 2010).

While the Thompson metric is convenient for solving many optimization problems involving matrices, it is often more intuitive to view matrices solving such problems within more typical geometric contexts. Knowing that the solution of a problem \( X \) and its \( n^{th} \) approximation \( X_n \) are \( d_T \) units apart in the Thompson metric provides little indication of how close \( X_n \) is to \( X \), i.e. knowing that \( X_n \leq aX \) and \( X \leq aX_n \) in the Löwner ordering (Baksalary & Pukelsheim, 1991), where \( a = e^\delta \), does not intuitively bound \( \| X - X_n \| \) for any of the usual matrix norms \( \| \cdot \| \). But it is \( \| X - X_n \| \) in a suitable matrix norm, not \( d_T \), or similar expressions relating \( X \) and \( X_n \) in the Löwner ordering, that provides insight as to the quality of an approximation \( X_n \).

In particular, considering the matrices \( X_n \) and \( X \) as linear operators on Euclidean vector spaces, the spectral norm, i.e. a Schatten p-norm with \( p = \infty \), of \( X - X_n \) is the relevant measure of how well \( X_n \) approximates \( X \). Considering these matrices as themselves vectors in a Euclidean space, then the relevant assessment of how well \( X_n \) approximates \( X \) is the Frobenius norm, i.e. a Schatten norm with \( p = 2 \), of \( X - X_n \). Therefore, it is useful to know an upper bound for the Schatten p-norm \( \| X - X_n \|_p \) given some minimal information about \( X \) (e.g. its norm) as well as the Thompson metric \( d = d_T(X, X_n) \).

For a cone with normality constant \( \delta \), in a Banach Space, the following inequality holds (Lemmens & Nussbaumer, 2012; Nussbaum, 1988):

\[
\| X - Y \|_p \leq (1 + 2\delta)(e^{d_T(X,Y)} - 1) \max(\| X \|_p, \| Y \|_p). 
\]

However, this inequality does not preclude the existence of tighter bounds relating specific norms and Thompson metrics such as Schatten p-norms and the Thompson metric induced by the Löwner order on the cone of positive semidefinite matrices.

This paper thus seeks to fill this important gap in our understanding of the relationship between Thompson metrics and Schatten norms by providing an upper bound for the Schatten p-norm \( \| X - Y \|_p \) given the Thompson metric \( d = d_T(X, Y) \) as well as the Schatten p-norms of \( X \) and \( Y \). In particular, the application of Weyl’s inequalities establishes that \( \| X - Y \|_p \leq 2\delta \frac{(e^{d} - 1)}{e^{d} - 1} \max(\| X \|_p, \| Y \|_p) \). Hopefully, this paper will serve as the beginning of a conversation leading to ever tighter bounds on \( \| X - Y \|_p \) given \( d = d_T(X, Y) \) as well as minimal information about \( X \) and \( Y \), such as their norms and perhaps some knowledge of their spectra of eigenvalues.
2. Preliminaries

This paper will generally use a consistent set of letters and symbols to denote certain matrices and their norms and eigenvalues. Let $X$ and $Y$ each denote positive definite Hermitian matrices with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and $\mu_1 \geq \cdots \geq \mu_n$, respectively. While the proofs presented in this paper do not explicitly require the matrices be positive definite, in such cases the Thompson metric may be infinite, when the matrix is not positive definite the results presented here are trivial as any finite metric is $\leq \infty$. Thus, this paper will focus on positive definite matrices $X$ and $Y$. Denote the eigenvalues of the matrix $\Delta = X - Y$ by $\delta_1 \geq \cdots \geq \delta_n$ and those of $E = -\Delta = Y - X$ by $\epsilon_1 \geq \cdots \geq \epsilon_n$. Note that $\delta_i = -\epsilon_{n-i+1}$. $\|M\|_p$ denotes a Schatten $p$-norm of the matrix $M$ and $\|M\|_p^{1/p}$ specifically denotes the Schatten $p$-norm (which is a norm for $p$ such that $1 \leq p \leq \infty$). Note that $\|M\|_p$ is a function of the eigenvalues $\mu_1 \geq \cdots \geq \mu_n$ of $M$: $\|M\|_p = f_p(\mu_1, \ldots, \mu_n) = \left(\sum_{i=1}^{n} |\mu_i|^p\right)^{1/p}$. Similarly, this paper will use the notation of $f(\mu_1, \ldots, \mu_n)$ as the functional form of $\|M\|_p$. Depending on the context, $\leq$ and $\geq$ denote either the usual ordering on real numbers or the Löwner ordering on matrices: i.e. $X \leq Y$ indicates that $Y - X$ is positive semidefinite. In terms of the Löwner ordering, the Thompson metric $d_T(X, Y)$ is the minimal $d_T = \log(\alpha)$ such that $Y \leq \alpha X$ and $X \leq \alpha Y$ (Nussbaum & Walsh, 2004). As is standard, $\text{tr}(M)$ denotes the trace of the matrix $M$.

Key to the proofs in this paper are the well-established Weyl’s inequalities (Bhatia, 2007; Weyl, 1949) for the eigenvalues of Hermitian matrices: let $M$, $Y$ and $P$ be Hermitian matrices such that $M = Y + P$. Denote the eigenvalues of $M$, $Y$ and $P$ by $\mu_1 \geq \cdots \geq \mu_n$, $\lambda_1 \geq \cdots \geq \lambda_n$, and $\rho_1 \geq \cdots \geq \rho_n$, respectively. Then, (Weyl’s inequalities) $\lambda_1 + \rho_n \leq \mu_1 \leq \lambda_1 + \rho_1$.

Use of Mathematica (Wolfram Research I, 2015) proved invaluable in simplifying the equations and inequalities presented in this paper. Numerical results were calculated using MATLAB (MathWorks I, 2017).

3. Proof of general case

The proof begins with a lemma applying Weyl’s inequalities to bound the eigenvalues of $\Delta = X - Y$ by the eigenvalues of $Y$ given upper and lower bounds for $X$ in the Löwner ordering. The second lemma, a consequence of the first lemma, bounds the eigenvalues of $\Delta = X - Y$ by the eigenvalues of $X$.

**Lemma 3.1**: Consider (positive definite) Hermitian matrices $X$ and $Y$ such that $X \leq \alpha Y$ and $X \geq \beta Y$. Then (A) $|\lambda_i| \leq |\alpha - 1| \cdot \lambda_i$ and (B) $|\delta_i| \leq \max(|\beta - 1|, |\alpha - 1|) \cdot \lambda_i$.

**Proof**:

Note that $\Delta = X - \alpha Y = (\alpha - 1) \cdot Y = X - \beta Y = (\beta - 1) \cdot Y$. Let $\lambda_1$ be the maximum eigenvalue of $X - \alpha Y$ (which is negative semi-definite as $\alpha Y$ is positive definite by the definition of $X \leq \alpha Y$) and $\lambda_1$ be the minimum eigenvalue of $X - \beta Y$ (which is positive semidefinite by the definition of $X \geq \beta Y$). By hypothesis, $\lambda_1 \leq 0$ and $\lambda_1 \geq 0$. Note that the eigenvalues for $(\alpha - 1) \cdot Y$ and $(\beta - 1) \cdot Y$ are $|\alpha - 1| \cdot \lambda_1$, $|\beta - 1| \cdot \lambda_1$, $|\alpha - 1| \cdot \lambda_n$, $|\beta - 1| \cdot \lambda_n$, respectively. By Weyl’s inequalities, we have $|\alpha - 1| \cdot \lambda_1 \geq \delta_1 \geq |\beta - 1| \cdot \lambda_1 \cdot \lambda_1$. Since $\lambda_1 \leq 0$ and $\delta_1 \geq 0$, we have $(\alpha - 1) \cdot \lambda_1 \geq \delta_1 \geq (\beta - 1) \cdot \lambda_1$ and hence $|\delta_i| \leq \max(|\beta - 1|, |\alpha - 1|) \cdot \lambda_1$.

**Lemma 3.2**: Again, consider (positive definite) Hermitian matrices $X$ and $Y$ such that $X \leq \alpha Y$ and $X \geq \beta Y$. Then (A) $|\delta_i| \leq \max(|\beta - 1|, |\alpha - 1|) \cdot \lambda_i$ and (B) $|\delta_i| \leq \max\left(\frac{1}{\beta}, \frac{1}{\alpha}\right) \cdot \lambda_i$.

**Proof**:

$X \leq \alpha Y$ and $X \geq \beta Y$ respectively imply $\frac{1}{\beta} \leq Y$ and $\frac{1}{\alpha} \geq X$. Apply Lemma 3.1 to $Y$ (in place of $X$), $X$ (in place of $Y$), $\frac{1}{\alpha}$ (in place of $\beta$) and $\frac{1}{\beta}$ (in place of $\alpha$).
Theorem 3.3: Consider (positive definite) Hermitian matrices $X$ and $Y$ such that $X \leq \alpha Y$ and $X \geq \beta Y$. \[\|X - Y\| \leq \min\left\{\max\{\|\alpha - 1\|, \|\beta - 1\|\}, \|Y\|, \max\left[\left\{\frac{1}{\alpha} - 1\right\}, \left\{\frac{1}{\beta} - 1\right\}\right] \cdot \|X\|\right\}\]

Proof:
Consider two sets of eigenvalues, $\lambda_1 \geq \ldots \geq \lambda_n$ and $\mu_1 \geq \ldots \geq \mu_n$ such that $\lambda_i \leq \mu_i$. Since $f$ is a functional form of a Schatten norm, it is monotonic in each variable, $\lambda_i \leq \mu_i$ implies $f(\lambda_1, \ldots, \lambda_n) \leq f(\mu_1, \ldots, \mu_n)$. Given that implication and that $f(\gamma_1, \ldots, \gamma_n) = f(\lambda_1, \ldots, \lambda_n)$, we have $|\delta| \leq \max\{\|\beta - 1\|, \|\alpha - 1\|\}$, which is given by part (B) of Lemma 3.1, implies that $f(\delta_1, \ldots, \delta_n) \leq \max\{\|\beta - 1\|, \|\alpha - 1\|\} \cdot \|X\|$. Similarly part (B) of Lemma 3.2 yields $|\delta| \leq \max\left[\left\{\frac{1}{\alpha} - 1\right\}, \left\{\frac{1}{\beta} - 1\right\}\right] \cdot \|X\|$, which implies that $f(\delta_1, \ldots, \delta_n) \leq \max\left[\left\{\frac{1}{\alpha} - 1\right\}, \left\{\frac{1}{\beta} - 1\right\}\right] \cdot \|X\|$. Combining these two inequalities for $f(\delta_1, \ldots, \delta_n)$ with the definition of $f(\mu_1, \ldots, \mu_n) = \|M\|$ yields $\|X - Y\| \leq \min\{\max\{\|\alpha - 1\|, \|\beta - 1\|\}, \|Y\|, \max\left[\left\{\frac{1}{\alpha} - 1\right\}, \left\{\frac{1}{\beta} - 1\right\}\right] \cdot \|X\|\}$.

Theorem 3.4: Consider (positive definite) Hermitian matrices $X$ and $Y$ such that Thompson metric $d = d_f(X, Y)$ is finite. Let $\lambda_1, \ldots, \lambda_n$ denote a collection of numbers such that $|\lambda_i| = \max\{\min|\lambda_i|, |\alpha_i|, \min|\lambda_{n-i+1}|, |\alpha_{n-i+1}|\} \cdot \frac{\alpha - 1}{\alpha}$. Then $\|X - Y\| \leq f(\lambda_1, \ldots, \lambda_n)$.

Proof:
Let $\lambda = e^{\frac{1}{c}}$. By definition of the Thompson metric, we have (i) $X \leq \alpha Y$ and (ii) $Y \leq \alpha X$. Let $\beta = \alpha^{-1} = e^{-d}$. Then we have (iii) $\delta X \leq Y$ and (iv) $\delta Y \leq X$. Applying Lemma 3.1, part A to (i) and (iii) yields $(\alpha - 1) \cdot \delta_i \geq \beta \cdot \delta_i \geq \delta_i \cdot (\beta - 1)$. Noting that $\beta = \frac{1}{\alpha}$ and $\alpha = \frac{1}{\beta}$, applying Lemma 3.2, part A to (i) and (iii) yields $\delta_i \cdot (\alpha - 1) \cdot \lambda_i \geq \delta_i \cdot (\beta - 1) \cdot \lambda_i$. Similarly, recall that $E = Y - X$, so reversing the roles of $X$ and $Y$ when applying Lemma 3.1, part A, and respectively Lemma 3.2, part A, to (ii) and (iv) yields $\delta_i \cdot (\alpha - 1) \cdot \lambda_i \geq \delta_i \cdot (\beta - 1) \cdot \lambda_i$ and $\delta_i \cdot (\alpha - 1) \cdot \lambda_i \leq \delta_i \cdot (\beta - 1) \cdot \lambda_i$. Since $\delta_i = e^{-n-i+1}$, application of Lemmas 3.1 and 3.2 (part A) to (ii) and (iv) yield $\delta_i \cdot (\alpha - 1) \cdot \lambda_i \cdot \lambda_{n-i+1} \leq \delta_i \cdot (\beta - 1) \cdot \lambda_i \cdot \lambda_{n-i+1}$. Since (by the definition of the Thompson metric) $\alpha \geq 1$ and hence $(\alpha - 1) \geq 1 - \beta = \frac{\alpha - 1}{\alpha}$, $(\beta - 1) \cdot \lambda_i \cdot \lambda_{n-i+1} \leq \delta_i \cdot (1 - \beta) \cdot \lambda_i \cdot \lambda_{n-i+1}$ and $(\beta - 1) \cdot \lambda_i \cdot \lambda_{n-i+1} \leq \delta_i \cdot (1 - \beta) \cdot \lambda_i \cdot \lambda_{n-i+1}$. Substituting $\frac{(1 - \beta)}{\alpha} = \frac{(1 - \beta)}{e}$, we have $|\delta| \leq \max\{\min|\lambda_i|, |\alpha_i|, \min|\lambda_{n-i+1}|, |\alpha_{n-i+1}|\} \cdot \frac{\alpha - 1}{\alpha} = |\lambda_i|$. As in the proof of Theorem 3.3, the monotonicity of $f$ and definition of $|\delta|$ yield $\|X - Y\| \leq f(\lambda_1, \ldots, \lambda_n)$.

Theorem 3.5: Consider (positive definite) Hermitian matrices $X$ and $Y$, i.e. such that Thompson metric $d = d_f(X, Y)$ is finite. Then $\|X - Y\| \leq 2\max\left[\frac{\|e^{-d}\|_p}{\max\|X\|_p, \|Y\|_p}\right]$. 

Proof:
Note that $\max\{\min|\lambda_i|, |\alpha_i|, \min|\lambda_{n-i+1}|, |\alpha_{n-i+1}|\} \leq \max\{\max|\lambda_i|, |\alpha_i|, \min|\lambda_{n-i+1}|, |\alpha_{n-i+1}|\}$

Since raising positive numbers to powers $\geq 1$ is monotonically increasing, a consequence of Theorem 3.4 is that $\delta_i \leq \frac{(e^{-d})_{\alpha} - 1}{e^{-d}}$. Since the maximum of $\delta_i$ is $\max\{\delta_i\}_{i=1}^{n} = \max\{\delta_i\}_{i=1}^{n}$, we have $\|X - Y\| \leq \max\{\sum_{i=1}^{n} |\lambda_i|^p + |\alpha_i|^p, |\lambda_{n-i+1}|^p + |\alpha_{n-i+1}|^p\}$. Hence, by the definition of and monotonicity of $f_p$, we have $\|X - Y\| \leq \max\{\sum_{i=1}^{n} |\lambda_i|^p + |\alpha_i|^p, |\lambda_{n-i+1}|^p + |\alpha_{n-i+1}|^p\}$.
definition of the Schatten p-norm yields \( \| X - Y \|_p^p \leq 2 \cdot \left( \frac{e^{p-1}}{e^p - 1} \right)^p \max \{ \| X \|_p^p, \| Y \|_p^p \} \). Taking the p'th root of both sides of the inequality yields the result.

4. The Frobenius (p = 2) case

We begin by noting that \( \text{tr}(A^TB) \) defines an inner product yielding the Frobenius norm, i.e. \( \| A \|_2 = \sqrt{\text{tr}(A^TA)} \). This, together with the commutative property of the trace, leads to the following version of the law of cosines for matrices: \( \| A - B \|_2^2 = \| A \|_2^2 + \| B \|_2^2 - 2 \cdot \text{tr}(A^TB) \). Since for two (symmetric) positive semidefinite matrices \( X \) and \( Y \), \( \text{tr} X^T Y = \text{tr} XY \) (Yang, 2000; Yang, Yang, & Teo, 2001), \( \theta = \cos^{-1} \left[ \frac{\text{tr}(XY)}{\sqrt{\| X \|_2^2 \| Y \|_2^2}} \right] \leq \frac{\pi}{2} \text{ (rad)} \) and hence \( \| X - Y \|_2^2 \leq \| X \|_2^2 + \| Y \|_2^2 \). Note that the Frobenius norm of a matrix is the same as the Euclidean norm of that matrix reshaped as a vector, so matrices under the Frobenius norm can be treated just as vectors in a Euclidean space.

Let \( d = d_T(X, Y) \) be the Thompson metric between (positive definite) matrices \( X \) and \( Y \) and let \( \alpha = e^d \). Note that in Figure 1, which represents each matrix as a vector, all the vectors shown are coplanar. The angle \( \theta \) is the same as the angle between \( \alpha X - Y = x + z \) and \( \alpha Y - X = y + w \). Since, by definition of the Thompson metric, \( \alpha X - Y \) and \( \alpha Y - X \) are both positive semidefinite, \( \theta \leq \frac{\pi}{2} \text{ (rad)} \) and \( \text{tr}(xy) \) and \( \text{tr}(wz) \geq 0 \) whereas \( \text{tr}(xw) \) and \( \text{tr}(yz) \leq 0 \). Thus, we have

\[
\| \Delta \|_2^2 \leq \| x \|_2^2 + \| y \|_2^2
\]

\[
\| \alpha \Delta \|_2^2 \leq \| w \|_2^2 + \| z \|_2^2
\]

\[
(\alpha - 1)X_2^2 \geq \| y \|_2^2 + \| z \|_2^2
\]

\[
(\alpha - 1)Y_2^2 \geq \| w \|_2^2 + \| x \|_2^2
\]

Adding inequalities (1) and (2) as well as (3) and (4), we have

\[
\| \Delta \|_2^2 + \| \alpha \Delta \|_2^2 \leq \| x \|_2^2 + \| y \|_2^2 + \| w \|_2^2 + \| z \|_2^2
\]

\[
(\alpha - 1)X_2^2 + (\alpha - 1)Y_2^2.
\]

Thus

\[
\| \Delta \|_2^2 + \| \alpha \Delta \|_2^2 \leq (\alpha - 1)X_2^2 + (\alpha - 1)Y_2^2.
\]
Solving for $\| \Delta \|_2$, we have our result

$$\| \Delta \|_2 \leq \frac{\alpha - 1}{\sqrt{\alpha + 1}} \sqrt{\| X \|_2^2 + \| Y \|_2^2}.$$  

(7)

5. A generalization of the Frobenius ($p = 2$) case

Consider the related and more general problem of bounding the Frobenius norm $\| \Delta \|_2$, with $\Delta = X - Y$, given matrix bounds $X \leq g \cdot Y$ and $Y \leq f \cdot X$, for scalars $f$ and $g$. This generalization, illustrated in Figure 2, yields the following equations:

$$\| \Delta \|_2^2 = \| X \|_2^2 + \| Y \|_2^2 - 2 \| X \|_2 \| Y \|_2 \cdot \cos \phi$$

(8)

$$\| \Delta_{fg} \|_2^2 = \| f \cdot X \|_2^2 + \| g \cdot Y \|_2^2 - 2 \| f \cdot X \|_2 \| g \cdot Y \|_2 \cdot \cos \phi = \ldots$$

(9)

$$f^2 \cdot \| X \|_2^2 + g^2 \cdot \| Y \|_2^2 - 2fg \| X \|_2 \| Y \|_2 \cdot \cos \phi.$$  

Similar to the argument above (in part 4), $X \leq g \cdot Y$ and $Y \leq f \cdot X$ imply that $\theta \leq \frac{\pi}{2} (\text{rad})$, which implies

$$\| \Delta \|_2^2 + g^2 \| Y \|_2^2 \leq (\| g \cdot Y \|_2 - \| Y \|_2)^2 + (\| f \cdot X \|_2 - \| X \|_2)^2 = \ldots$$

(10)

$$\| f \cdot X \|_2^2 + (g - 1)^2 \cdot \| Y \|_2^2.$$  

The above system of two quadratic equations and one quadratic inequality has (assuming $\| X \|_2$ is known, even if $X$ is an unknown, approximated by $Y$) three unknowns: $\| \Delta \|_2$, $\| \Delta_{fg} \|_2$ and $\cos \phi$. Solving this system and simplifying the resulting solutions with Mathematica (Wolfram Research Inc., 2016) yields the following inequalities:

$$\| \Delta \|_2 \leq \sqrt{\frac{(1 + f(f - 2)) \cdot \| X \|_2^2 + (1 + g(f - 2)) \cdot \| Y \|_2^2}{1 + f \cdot g}}$$

(11)

$$\| \Delta_{fg} \|_2 \leq \sqrt{\frac{f^2 \cdot ((f - 2) \cdot g + 1)) \cdot \| X \|_2^2 + (g^2 \cdot ((g - 2) \cdot f + 1)) \cdot \| Y \|_2^2}{1 + f \cdot g}}$$

(12)

Figure 2. Difference and Frobenius norm between two vectors give matrix bounds. This figure represents matrices $X$ and $Y$ as vectors that span a plane and illustrates the geometric intuition behind Equations (8) and (9) as well as inequality (10).
\[ \cos \phi \geq \frac{f \cdot \| X \|_2^2 + g \cdot \| Y \|_2^2}{1 + f \cdot g \cdot \| X \|_2 \cdot \| Y \|_2}. \]  

(13)

Note that this not only establishes a bound for \( \| X - Y \|_2 \), given matrix bounds \( X \leq g \cdot Y \) and \( Y \leq f \cdot \sqrt{b^2 - 4ac} \cdot X \), but this analysis also yields a bound for \( \cos \phi \). Thus, this analysis provides information about the inner product between \( X \) and \( Y \), even in cases where \( X \) is an unknown, approximated by \( Y \). Of course, when \( f = g = \alpha = e^d \), (11) simplifies to the result established in Section 4.

6. Numerical results

Fifty calculations, performed in MATLAB, with pairs of random positive definite \( 5 \times 5 \) matrices tested the tightness of the bounds presented in this paper. The following formulas generated the \( i^{th} \) pair \( (X_i, Y_i) \) of matrices

\[ X_i = (A + (i^2/100) \cdot B) (A + (i^2/100) \cdot B)^T + D \]  

(14)

\[ Y_i = (A + (i^2/100) \cdot C) (A + (i^2/100) \cdot C)^T + D, \]  

(15)

where \( A, B \) and \( C \) have elements randomly drawn from the uniform distribution on \([0,1]\) and \( D \) is a diagonal matrix with diagonal elements randomly drawn from that same distribution. The occurrence of the index \( i \) in the formula ensured a range of distances among the 50 matrix pairs tested.

Figure 3 compares values of \( X_i - Y_i \) for (A) \( p = 1 \) (trace norm), (B) \( p = 2 \) (Frobenius norm) and (C) \( p = \infty \) (spectral norm) with bounds for those values calculated using Theorem 3.5, (and for panel B) Equation (7) and Equation (11). The function, thompson_metric.m, used to calculate the Thompson metric as well as \( f \) and \( g \) in Equation (11) is available via MATLAB Central File Exchange, and the script, and data calculated using that script, used to generate Figure 3 is available from the author upon request. While the bounds described in this paper are clearly not very tight (for matrices more distant from each other), hopefully, these results will spark further research leading to tighter bounds on Schatten norms based on the Thompson metric.
7. Discussion

Weyl’s inequalities, and hence some knowledge of the spectra of \( X \) and \( Y \), form the backbone of the proofs presented above. In the motivating case where \( Y \) is an approximation of an unknown \( X \), the spectrum of \( Y \) may also be unknown. While the principle result of this paper ultimately only requires knowledge of \( \| X \|_p \) (as well as \( \| Y \|_p \), which is generally known), purely geometric/trigonometric proofs, such as the one given for the Frobenius case, of the results presented in this paper would be more elegant given the nature of the motivating problem.

Furthermore, proofs not based on the matrix structure of \( X \) and \( Y \) but based purely on the ordering (Löwner ordering in this case) and norm (Schatten \( p \)-norm) being compared might allow for tighter bounds on \( \| X - Y \|_p \) even in the absence of any knowledge of the spectrum of \( X \) (or even of \( Y \), for that matter), other than perhaps a restriction that \( X \) and \( Y \) be positive semidefinite. In comparison, Theorem 3.4 provides a tighter bound on \( \| X - Y \|_p \) than the main result (Theorem 3.5), but it requires some knowledge of the spectrum of \( X \) (at least that its eigenvalues are lower in magnitude than the corresponding eigenvalues of \( Y \)).

Additionally, proofs not based on the matrix structure of \( X \) and \( Y \) may lead to the generalization of these results in other orderings, which can also induce Thompson metrics (Cobzas & Rus, 2014), and other norms. For instance, since the Frobenius norm arises from an inner product, a geometrically flavored argument leads to a slightly tighter bound on \( \| X - Y \|_2 \) than obtained from the general bound for \( \| X - Y \|_p \) and setting \( p = 2 \). On the other hand, the already established general result for a Thompson metric induced by a normal cone in a Banach space (Lemmens & Nussbaum, 2012; Nussbaum, 1988) is not as tight as the main result (Theorem 3.5) presented here: as \( \epsilon \to 0 \), the value of \( \delta \) such that \( 0 \leq X \leq (X + I) \Rightarrow \| X \| \leq \delta \| X + I \| \) approaches unity; thus, the normality constant for the cone of positive semidefinite matrices is unity, and the general result for Banach spaces reduces to \( \| X - Y \| \leq 3 (e^{\delta d(X,Y)} - 1) \max \{\| X \|, \| Y \|\} \), the right-hand side of which inequality is clearly greater than \( 2^{1/2} \max \{\| X \|_p, \| Y \|_p\} \) since \( 2^{1/2} \leq 2 < 3 \) and \( e^\delta \geq 1 \) since the (Thompson) metric \( d \) is non-negative.

As illustrated in Section 5 of this paper, more general analysis of the Frobenius case yields not only a bound for \( \| X - Y \|_2 \) but also bounds inner product between \( X \) and \( Y \). In the case where \( X \) is an unknown, approximated by \( Y \), bounds on the inner product between \( X \) and \( Y \) further quantify how well \( Y \) approximates \( X \), and may provide further insight into improved approximations of an unknown \( X \).

Hopefully, future research can further generalize the analysis presented in Section 5 to cases where \( X \leq g(Y) \) and \( Y \leq f(X) \), for more general classes of functions on \( X \) and \( Y \) than mere scalar multiplication. Such inequalities, in the Löwner order, arise, for example, in characterizing approximate solutions to the continuous algebraic Riccati equation (Zhang & Liu, 2010). Further generalization of the results presented here will facilitate expressing the quality of approximations, found in many areas of matrix algebra and optimization theory, in terms of geometrically intuitive metrics based on Schatten norms rather than less geometrically intuitive bounds in the Löwner order.

Acknowledgements

This work was made possible by the support of William Paterson University of NJ’s Office of the Provost for Assigned Release Time for research. The author also thanks Rajendra Bhatia as well as Yongdo Lim for their advice about where to submit this work and an anonymous reviewer for referring me to Lemmens’ and Nussbaum’s related results in nonlinear Perron-Frobenius theory.

Funding

This work was not funded by any external granting agency or funding source. The author thanks the Department of Chemistry at William Paterson University for providing funding to maintain the author’s MATLAB license.

Author details

David A. Snyder
E-mail: snyderd@wpunj.edu
ORCID ID: http://orcid.org/0000-0001-6608-2975
Department of Chemistry, College of Science and Health, William Paterson University of NJ, 300 Pompton Road, Wayne, NJ 07470, USA.

Disclosure Statement

The author has no competing interests to declare.
Citation information

Cite this article as: On bounding the Thompson metric by Schatten norms, David A. Snyder, Cogent Mathematics & Statistics (2019), 6: 1614318.

References


