Norm inequalities

1 Trace norm and fidelity[1]

Definition 1. The *trace norm* of an operator $A \in \mathrm{L}(\mathcal{H})$ is $\|A\|_1 = \mathrm{Tr}\left(\sqrt{A^\dagger A}\right) = \sum_j \lambda_j$, where λ_j are the singular values of A. (For Hermitian operators, these are just the absolute values of the eigenvalues.)

Proposition 2. *The function $\|\cdot\|_1$ satisfies the axioms of a norm. For any matrices A and B, this inequality holds:*

$$|\mathrm{Tr}\, AB| \leqslant \|A\|_1 \|B\|. \tag{1}$$

Furthermore,

$$\boxed{\|A\|_1 = \max_{B\colon \|B\|\leqslant 1} |\mathrm{Tr}\, AB| = \max_{U\in \mathrm{U}(\mathcal{H})} |\mathrm{Tr}\, AU|.} \tag{2}$$

Proof. Let $A = \sum_j \lambda_j |\psi_j\rangle\langle\nu_j|$ be a singular value decomposition of A, where $\lambda_j > 0$ and$\langle\psi_j|\psi_k\rangle = \langle\nu_j|\nu_k\rangle = \delta_{jk}$. Then,

$$|\mathrm{Tr}\, AB| = \left|\sum_j \lambda_j \langle\nu_j|B|\psi_j\rangle\right| \leq \sum_j \lambda_j \big|\langle\nu_j|B|\psi_j\rangle\big| \leq \sum_j \lambda_j \|B\| = \|A\|_1 \|B\|$$

for any B. Hence,

$$\|A\|_1 \geqslant \max_{B\colon \|B\|\leqslant 1} |\mathrm{Tr}\, AB| \geqslant \max_{U\in \mathrm{U}(\mathcal{H})} |\mathrm{Tr}\, AU|.$$

On the other hand, if U is a unitary operator that takes $|\psi_j\rangle$ to $|\nu_j\rangle$, then $\mathrm{Tr}\, AU = \|A\|_1$.

We now show that $\|\cdot\|_1$ is a norm. The only nontrivial property is the triangle inequality. It can be derived as follows:

$$\begin{aligned}\|A_1 + A_2\|_1 &= \max_{B\colon \|B\|\leqslant 1} \big|\mathrm{Tr}(A_1 + A_2)B\big| \\ &\leq \max_{B\colon \|B\|\leqslant 1} |\mathrm{Tr}\, A_1 B| + \max_{B\colon \|B\|\leqslant 1} |\mathrm{Tr}\, A_2 B| = \|A_1\|_1 + \|A_2\|_1.\end{aligned}$$

□

Equations of the form $\|A\|_1 = \max_{B\colon \|B\|\leqslant 1} |\mathrm{Tr}\, AB|$ are called *norm dualities*. This particular equation says that the trace norm is dual to the operator norm. The converse is also true, i.e. $\|B\| = \max_{A\colon \|A\|_1\leqslant 1} |\mathrm{Tr}\, AB|$.

Proposition 3. *The trace norm has the following properties:*

a) $\|AB\|_1, \|BA\|_1 \leq \|A\|_1 \|B\|$,   *b)* $|\mathrm{Tr}\, A| \leq \|A\|_1$,
c) $\|\mathrm{Tr}_{\mathcal{M}}\, A\|_1 \leq \|A\|_1$,   *d)* $\|A \otimes B\|_1 = \|A\|_1 \|B\|_1$.

[1] Adapted from Kitaev, Shen, Vyalyi, "Classical and quantum computation".

Proof. Property a):

$$\|AB\|_1 = \max_{U\in \mathrm{U}(\mathcal{H})} |\operatorname{Tr} ABU| \le \max_{U\in \mathrm{U}(\mathcal{H})} \|A\|_1\|BU\| = \|A\|_1\|B\|.$$

Property b) is a special case of c), so we prove c). Let $A \in \mathrm{L}(\mathcal{H}\otimes\mathcal{M})$; then

$$\|\operatorname{Tr}_{\mathcal{M}} A\|_1 = \max_{U\in \mathrm{U}(\mathcal{H})} \bigl|\operatorname{Tr}\bigl((\operatorname{Tr}_{\mathcal{M}} A)U\bigr)\bigr| = \max_{U\in \mathrm{U}(\mathcal{H})} \bigl|\operatorname{Tr}\bigl(A(U\otimes I_{\mathcal{M}})\bigr)\bigr| \le \|A\|_1.$$

Property d): If λ_j and μ_k are the singular values of A and B, respectively, then the operator $A\otimes B$ has singular values $\lambda_j\mu_k$. □

The *trace norm distance* between density matrices ρ and γ is defined as $\frac{1}{2}\|\rho-\gamma\|_1$. This quantity vanishes if $\rho=\gamma$ and achieves the maximum equal to 1 if they are supported by orthogonal subspaces. The trace norm distance is compatible with the Kolmogorov distance between probability distributions, $\frac{1}{2}\sum_j |u_j - v_j|$. More exactly, if we perform the same measurement given by a set of orthogonal projectors Π_j on states ρ and γ, the Kolmogorov distance between the resulting probability distributions is bounded by the distance between the original states:

$$\sum_j \bigl|\operatorname{Tr}(\rho\Pi_j) - \operatorname{Tr}(\gamma\Pi_j)\bigr| \le \|\rho-\gamma\|_1. \tag{3}$$

Indeed, the left-hand side can be represented in the form $\operatorname{Tr}((\rho-\gamma)U)$, where $U=\sum_j \tau_j\Pi_j$ with $\tau_j = \operatorname{sgn}\operatorname{Tr}((\rho-\gamma)\Pi_j) = \pm 1$. It is clear that U is unitary, hence $|\operatorname{Tr}((\rho-\gamma)U)| \leqslant \|\rho-\gamma\|_1$.

Another commonly used distance measure for quantum states is *fidelity*:[2]

$$F(\rho,\gamma) \;:=\; \max\Bigl\{\bigl|\langle\xi|\eta\rangle\bigr|^2 \;:\; \operatorname{Tr}_{\mathcal{F}}(|\xi\rangle\langle\xi|)=\rho,\;\; \operatorname{Tr}_{\mathcal{F}}(|\eta\rangle\langle\eta|)=\gamma\Bigr\}. \tag{4}$$

Here $\rho,\gamma\in\mathrm{L}(\mathcal{H})$ are density matrices, and $\mathcal{F}$ is some auxiliary Hilbert space of sufficient dimension to allow the purification, i.e. $\dim\mathcal{F} \ge \max\{\operatorname{rank}(\rho),\operatorname{rank}(\gamma)\}$. Note that F has the opposite behavior of what one usually calls a distance: it has a maximum value (equal to 1) when ρ and γ coincide and vanishes when they are supported by orthogonal subspaces.

One useful formula is

$$\boxed{F(\rho,\gamma) = \bigl\|\sqrt{\rho}\sqrt{\gamma}\bigr\|_1^2.} \tag{5}$$

To show this, let $\mathcal{F}=\mathcal{H}^*$ so that the vectors $|\xi\rangle,|\eta\rangle\in\mathcal{H}\otimes\mathcal{H}^*$ can be associated with operators $X,Y\in\mathrm{L}(\mathcal{H})$ (due to the isomorphism $\mathcal{H}\otimes\mathcal{H}^*\cong\mathrm{L}(\mathcal{H})$). For example, if $\xi=\sum_j \lambda_j\,\psi_j\otimes\mu_j$, then $X=\sum_j\lambda_j\,\psi_j\mu_j$, where in the first case $\mu_j\in\mathcal{H}^*$ could be written as a ket-vector, and in the second as a bra-vector. Thus, the condition $\operatorname{Tr}_{\mathcal{F}}(|\xi\rangle\langle\xi|)=\rho$ becomes $XX^\dagger=\rho$. One solution to this equation is $X=\sqrt{\rho}$; the most general solution is $X=\sqrt{\rho}\,U$, where U is an arbitrary unitary operator. Similarly, $Y=\sqrt{\gamma}\,V$, where V is unitary. Thus,

$$\langle\xi|\eta\rangle = \operatorname{Tr}(X^\dagger Y) = \operatorname{Tr}\bigl(U^\dagger\sqrt{\rho}\sqrt{\gamma}\,V\bigr) = \operatorname{Tr}\bigl(\sqrt{\rho}\sqrt{\gamma}\,W\bigr), \qquad \text{where } W=VU^\dagger,$$
$$F(\rho,\gamma) = \max_{W\in\mathrm{U}(\mathcal{H})}\bigl|\operatorname{Tr}\bigl(\sqrt{\rho}\sqrt{\gamma}W\bigr)\bigr|^2 = \bigl\|\sqrt{\rho}\sqrt{\gamma}\bigr\|_1^2.$$

[2]Preskill used the same notation in his lecture notes, whereas Nilson and Chuang's F is the square root of ours.

The fidelity and the trace norm are related by the following inequalities:

$$\boxed{\left(1-\frac{\|\rho-\gamma\|_1}{2}\right)^2 \;\le\; \left(\mathrm{Tr}\big(\sqrt{\rho}\sqrt{\gamma}\big)\right)^2 \;\le\; F(\rho,\gamma) \;\le\; 1-\left(\frac{\|\rho-\gamma\|_1}{2}\right)^2.} \tag{6}$$

To obtain the first of them, we will use the following lemma.

Lemma 4. *Let X and Y be nonnegative Hermitian operators. Then*

$$\mathrm{Tr}(X-Y)^2 \le \|X^2-Y^2\|_1.$$

Proof. Let $|\psi_j\rangle$, λ_j be orthonormal eigenvectors and the corresponding eigenvalues of the operator $X-Y$. We have the following bound:

$$\|X^2-Y^2\|_1 \ge \sum_j \big|\langle\psi_j|(X^2-Y^2)|\psi_j\rangle\big|.$$

(Indeed, the right-hand side can be represented as $\mathrm{Tr}((X^2-Y^2)U)$, where $U=\sum_j \pm|\psi_j\rangle\langle\psi_j|$.) To proceed, we need to estimate each term in the sum,

$$\begin{aligned}
&\langle\psi_j|(X^2-Y^2)|\psi_j\rangle\\
&\qquad=\frac{1}{2}\langle\psi_j|(X-Y)(X+Y)|\psi_j\rangle+\frac{1}{2}\langle\psi_j|(X+Y)(X-Y)|\psi_j\rangle\\
&\qquad=\lambda_j\langle\psi_j|(X+Y)|\psi_j\rangle,\\
&\langle\psi_j|(X+Y)|\psi_j\rangle \ge |\lambda_j|.
\end{aligned}$$

Thus,

$$\sum_j \big|\langle\psi_j|(X^2-Y^2)|\psi_j\rangle\big| \ge \sum_j \lambda_j^2 = \mathrm{Tr}(X-Y)^2.$$

□

Proof of the inequalities (6). By the previous lemma,

$$\mathrm{Tr}\big(\sqrt{\rho}\sqrt{\gamma}\big)=1-\frac{1}{2}\,\mathrm{Tr}\big(\sqrt{\rho}-\sqrt{\gamma}\big)^2 \ge 1-\frac{\|\rho-\gamma\|_1}{2}.$$

The inequality $(\mathrm{Tr}(\sqrt{\rho}\sqrt{\gamma}))^2 \le F(\rho,\gamma)$ follows from the fact that $\sqrt{\rho}$ and $\sqrt{\gamma}$ are particular purifications of ρ and γ, whereas the fidelity is the maximum over all purifications.

Let $|\xi\rangle$ and $|\eta\rangle$ provide the maximum in (4). Then

$$\begin{aligned}
\|\rho-\gamma\|_1 &= \big\|\mathrm{Tr}_{\mathcal{F}}\big(|\xi\rangle\langle\xi|-|\eta\rangle\langle\eta|\big)\big\|_1\\
&\le \big\||\xi\rangle\langle\xi|-|\eta\rangle\langle\eta|\big\|_1 = 2\sqrt{1-\big|\langle\xi|\eta\rangle\big|^2}=2\sqrt{1-F(\rho,\gamma)}.
\end{aligned}$$

Thus $F(\rho,\gamma)\le 1-\frac{1}{4}\|\rho-\gamma\|_1^2$, which is the required upper bound for the fidelity. □

2 Schatten norms

Definition 5. Let $1 \leqslant p < \infty$. The ℓ_p *norm* of a sequence of numbers $a = (a_1, \ldots, a_n)$ is

$$\|a\|_p = \left(\sum_{j=1}^{n} |a_j|^p \right)^{1/p}. \tag{7}$$

The *Schatten p-norm* of an operator $A \in \mathrm{L}(\mathcal{H})$ is $\|A\|_p = \|\lambda\|_p$, where $\lambda = (\lambda_1, \ldots, \lambda_n)$ is formed by the singular values of A. Furthermore,

$$\|a\|_\infty = \max_{j=1}^{n} |a_j|, \qquad \|A\|_\infty = \|A\|. \tag{8}$$

Note that $\|UAV\|_p = \|A\|_p$ for all unitary operators U and V.

Lemma 6 (Classical Hölder's inequality). *Let $a = (a_1, \ldots, a_n)$ and $b = (a_1, \ldots, b_n)$. Then*

$$\left| \sum_{j=1}^{n} a_j b_j \right| \leqslant \|a\|_p \, \|b\|_q, \qquad \text{where} \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{9}$$

Proof. Let us exclude the straightforward cases $p = \infty$ and $q = \infty$, and assume that both p and q are finite. It is clear that

$$\frac{\left| \sum_j a_j b_j \right|}{\|a\|_p \, \|b\|_q} \leqslant \sum_j \tilde{a}_j \tilde{b}_j, \qquad \text{where} \quad \tilde{a}_j = \frac{|a_j|}{\|a_j\|_p}, \quad \tilde{b}_j = \frac{|b_j|}{\|b_j\|_q}.$$

Note that $\|\tilde{a}\|_p = \|\tilde{b}\|_q = 1$. To show that $\sum_j \tilde{a}_j \tilde{b}_j \leqslant 1$, we will use Jensen's inequality. It states that if $w_j \geqslant 0$ and $\sum_j w_j = 1$, then

$$f\left(\sum_j w_j x_j \right) \leqslant \sum_j w_j f(x_j) \quad \text{if } f \text{ is convex,} \tag{10}$$

$$f\left(\sum_j w_j x_j \right) \geqslant \sum_j w_j f(x_j) \quad \text{if } f \text{ is concave} \tag{11}$$

(which follows immediately from the definition of a convex or concave function). Let us apply the second variant to the concave function $f(x) = x^{1/q}$ and

$$w_j = \tilde{a}_j^p, \qquad x_j = \frac{\tilde{b}_j^q}{\tilde{a}_j^p}.$$

The weights w_j satisfy the required conditions because $\sum_j w_j = \|\tilde{a}\|_p^p = 1$. Thus,

$$\sum_j \tilde{a}_j \tilde{b}_j = \sum_j w_j x_j^{1/q} \leqslant \left(\sum_j w_j x_j \right)^{1/q} = \|\tilde{b}\|_q = 1,$$

where we have used the fact that $1/p + 1/q = 1$, and hence, $p/q = p - 1$. $\square$

The operator analogue of Hölder's inequality reads

$$\boxed{\left|\operatorname{Tr} AB\right| \leqslant \|A\|_p \, \|B\|_q \qquad \text{if} \quad \frac{1}{p}+\frac{1}{q}=1.} \tag{12}$$

It follows from its classical version and the next theorem, which is useful on its own right.

Theorem 7 (von Neumann's trace inequality). *Let A and B be $n\times n$ matrices with singular values $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ and $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n \geqslant 0$. Then $|\operatorname{Tr} AB| \leqslant \sum_{k=1}^{n} \lambda_k\mu_k$.*

Proof. Let us consider the corresponding singular value decompositions,

$$A = \sum_j \lambda_j |\psi_j\rangle\langle\nu_j|, \qquad B = \sum_k \mu_k |\xi_k\rangle\langle\eta_k|.$$

Then

$$\operatorname{Tr} AB \;=\; \sum_{j,k} \lambda_j U_{jk}\,\mu_k V_{kj} \;=\; \sum_{j,k} \Bigl(\sqrt{\lambda_j}\,U_{jk}\sqrt{\mu_k}\Bigr)\Bigl(\sqrt{\mu_k}\,V_{kj}\sqrt{\lambda_j}\Bigr),$$

where $U_{jk} = \langle\nu_j|\xi_k\rangle$ and $V_{kj} = \langle\eta_k|\psi_j\rangle$. Clearly, U and V are unitary matrices. Now, let us use the Cauchy-Schwarz inequality:

$$\left|\operatorname{Tr} AB\right|^2 \;\leqslant\; \Biggl(\sum_{j,k} \lambda_j |U_{jk}|^2 \mu_k\Biggr)\Biggl(\sum_{j,k} \mu_k |V_{kj}|^2 \lambda_j\Biggr).$$

The numbers $|U_{jk}|^2$ and $|V_{kj}|^2$ form doubly stochastic matrices, i.e. their elements are nonnegative, and each row and column sums to 1. The next lemma says that in this situation, both sums in the above inequality are bounded by $\sum_k \lambda_k\mu_k$, which implies the theorem. □

Lemma 8. *Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ and $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n \geqslant 0$, and let W be a doubly stochastic $n\times n$ matrix. Then $\sum_{j,k} \lambda_j W_{jk}\mu_k \leqslant \sum_k \lambda_k\mu_k$.*

Proof. We first assume that W is a permutation matrix. In this special case, the lemma can be reformulated as follows. Let $\tilde\lambda_1,\dots,\tilde\lambda_n$ be the numbers $\lambda_1,\dots,\lambda_n$ listed in an arbitrary order; then $\sum_k \tilde\lambda_k\mu_k \leqslant \sum_k \lambda_k\mu_k$. To see this, we sort the $\tilde\lambda$'s by swapping pairs of numbers that are listed in the wrong order. In this process, the sum $\sum_k \tilde\lambda_k\mu_k$ may increase but not decrease. Indeed, if we swap the numbers in places $j<k$, then $\tilde\lambda_j \leqslant \tilde\lambda_k$ but $\mu_j \geqslant \mu_k$, hence the sum increases by the amount

$$(\tilde\lambda_k\mu_j + \tilde\lambda_j\mu_k) - (\tilde\lambda_j\mu_j + \tilde\lambda_k\mu_k) = (\tilde\lambda_k - \tilde\lambda_j)(\mu_j - \mu_k) \geqslant 0.$$

The general case follows from the Birkhoff-von Neumann theorem, which states that any doubly stochastic matrix is a convex combination of permutation matrices. A more elementary argument uses the induction on the number m of nonzero off-diagonal elements. Suppose that $m>0$ and that the lemma holds for all doubly stochastic matrices with less than m nonzero off-diagonal elements. Let us consider the graph with vertices $1,\dots,n$ that has an edge from k to j iff $j\neq k$ and $W_{jk}\neq 0$. Due to the double stochasticity, if a vertex v has an incoming edge, then $W_{vv}<1$, and hence, there is also an outgoing edge. Thus, the graph has a cycle $(v_1,\dots,v_l)$, which may be interpreted as a permutation:

$$\sigma(v_s) = v_{s+1} \;\text{ for }\; s=1,\dots,l-1, \quad \sigma(v_l)=v_1, \quad \text{and }\; \sigma(k)=k \;\text{ for other vertices } k.$$

Let us also denote by w the smallest of the numbers $W_{\sigma(k),k}$ for $k \in \{v_1, \dots, v_l\}$. Now, we construct a new doubly stochastic matrix by shifting some weight from the cycle edges to the diagonal, removing at least one edge completely:

$$\widetilde{W}_{jk} = W_{jk} + w\big(\delta_{jk} - \delta_{j,\sigma(k)}\big).$$

Since $\widetilde{W}$ has fewer nonzero off-diagonal elements than W, we have $\sum_{j,k} \lambda_j \widetilde{W}_{jk} \mu_k \leqslant \sum_k \lambda_k \mu_k$ by the induction hypothesis. On the other hand,

$$\sum_{j,k} \lambda_j \widetilde{W}_{jk} \mu_k - \sum_{j,k} \lambda_j W_{jk} \mu_k = w\left(\sum_k \lambda_k \mu_k - \sum_k \lambda_{\sigma(k)} \mu_k\right) \geqslant 0.$$

(This is the special case we considered first.) □

Proposition 9. *The function $\|\cdot\|_p$ satisfies the axioms of a norm. Furthermore, the norms corresponding to complementary pairs of p and q are dual to each other:*

$$\|A\|_p = \max\big\{|\operatorname{Tr} AB| : \|B\|_q \leqslant 1\big\}, \qquad \textit{where} \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{13}$$

The proof is analogous to that of Proposition 2.

Theorem 10 (Generalized Hölder's inequality for operators).

$$\boxed{\|AB\|_r \leqslant \|A\|_p \, \|B\|_q \qquad \textit{where} \quad \frac{1}{r} = \frac{1}{p} + \frac{1}{q}, \quad p, q, r \geqslant 1.} \tag{14}$$

Proof. One easy special case is $r = 2$:

$$\|AB\|_2^2 = \operatorname{Tr}\big((AB)^\dagger (AB)\big) = \operatorname{Tr}\big(A^\dagger A B B^\dagger\big) \leqslant \|AA^\dagger\|_{p/2} \|BB^\dagger\|_{q/2} = \|A\|_p^2 \|B\|_q^2,$$

where we have used the usual Hölder inequality for operators, i.e. Eq. (12). (The equation $\|AA^\dagger\|_{p/2} = \|A\|_p^2$ follows directly from the definition.)

We now consider the general case. By Eq. (13),

$$\|AB\|_r = \max\big\{|\operatorname{Tr} ABC| : \|C\|_s \leqslant 1\big\}, \qquad \text{where} \quad \frac{1}{s} + \frac{1}{r} = 1.$$

It remains to show that

$$|\operatorname{Tr} ABC| \leqslant \|A\|_p \|B\|_q \|C\|_s \qquad \text{if} \quad \frac{1}{p} + \frac{1}{q} + \frac{1}{s} = 1.$$

At least two of the three numbers $1/p$, $1/q$, $1/s$ are less than or equal to $1/2$. We may assume without loss of generality that these are $1/p$ and $1/q$. Therefore, there exist u and v such that

$$u, v \geqslant 0, \quad u + v = 1, \qquad \frac{u}{s} + \frac{1}{p} = \frac{v}{s} + \frac{1}{q} = \frac{1}{2}.$$

By exchanging some unitary factors between B and C and between C and A, we can arrange that $C = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Then

$$\begin{aligned} |\operatorname{Tr} ABC| &= \big|\operatorname{Tr} C^u A B C^v\big| \\ &\leqslant \|C^u A\|_2 \|B C^v\|_2 && \text{(Cauchy-Schwarz)} \\ &\leqslant \|C^u\|_{s/u} \|A\|_p \|B\|_q \|C^v\|_{s/v} && \text{(the } r = 2 \text{ case)} \\ &= \|C\|_s^u \|A\|_p \|B\|_q \|C\|_s^v = \|A\|_p \|B\|_q \|C\|_s. \end{aligned}$$

□

3 An application

The following inequality provides an upper bound for the trace norm distance between a fixed quantum state γ and another state ρ, which may depend on random parameters. This bound is given by an expression that is quadratic in ρ, and therefore, is easy to average.

$$\boxed{\|\rho-\gamma\|_1^2 \leqslant \mathrm{Tr}(\gamma^{-1/4}\rho\gamma^{-1/4})^2 - 1.} \tag{15}$$

To derive this inequality, we write

$$\rho-\gamma = ABA, \qquad A=\gamma^{1/4}, \qquad B=\gamma^{-1/4}\rho\,\gamma^{-1/4}-\gamma^{1/2}.$$

Using the generalized Hölder inequality twice, we get

$$\|ABA\|_1 = \|A\|_4\|B\|_2\|A\|_4 \qquad \text{because} \quad \frac{1}{4}+\frac{1}{2}+\frac{1}{4}=1.$$

Thus,

$$\begin{aligned}\|\rho-\gamma\|_1^2 \leqslant \|A\|_4^2\|B\|_2^2\|A\|_4^2 &= (\mathrm{Tr}\,\gamma)^{1/2}\Big(\mathrm{Tr}\big(\gamma^{-1/4}\rho\,\gamma^{-1/4}-\gamma^{1/2}\big)^2\Big)(\mathrm{Tr}\,\gamma)^{1/2}\\ &= \mathrm{Tr}(\gamma^{-1/4}\rho\,\gamma^{-1/4})^2 - 2\,\mathrm{Tr}\,\rho + \mathrm{Tr}\,\gamma = \mathrm{Tr}(\gamma^{-1/4}\rho\,\gamma^{-1/4})^2 - 1.\end{aligned}$$