AIMS Mathematics, 11(7): 23437–23449.
DOI: 10.3934/math.2026945
Received: 14 May 2026
Revised: 19 July 2026
Accepted: 24 July 2026
Published: 31 July 2026

https://www.aimspress.com/journal/Math

Research article

An affine matrix scalarization of the Cramér–Rao inequality

Songjun Lv*

School of Mathematics and Statistics, Suzhou University of Technology, Changshu, 215500, China

* **Correspondence:** Email: lvsongjun@126.com.

Abstract: We study an affine scalarization of the quadratic Cramér–Rao inequality. Instead of evaluating the usual trace product of the covariance matrix and the Fisher information matrix in a fixed Euclidean position, we minimize this product over all nonsingular linear positions. Using the positive-definite specialization of Alberti's variational characterization of matrix fidelity, while giving a self-contained proof that records all minimizers, we obtain, for positive definite matrices S and T,

$$\inf_{A\in \mathrm{GL}(n)} \operatorname{tr}(ASA^{\mathsf{T}})^{1/2} \operatorname{tr}(A^{-\mathsf{T}}TA^{-1})^{1/2} = \operatorname{tr}\left[\left(S^{1/2}TS^{1/2}\right)^{1/2}\right].$$

The matrix identity itself is classical; our contribution is its interpretation as an affine scalarization of the covariance–Fisher-information pair, including the optimal affine-position condition and the statistical equality cases. Applied to random vectors, the resulting scalar Cramér–Rao quantity is affine invariant, is no larger than the usual trace product in any fixed Euclidean position, and, when centered at the mean, attains the lower bound for every nonsingular Gaussian distribution.

Keywords: Cramér–Rao inequality; Fisher information; covariance matrix; affine scalarization; trace inequalities; matrix optimization; Gaussian extremizers
Mathematics Subject Classification: 62B10, 94A17

1. Introduction

The Cramér–Rao inequality is one of the fundamental principles relating error and information. In its classical form, it gives a lower bound for the covariance of an unbiased estimator in terms of the inverse of the Fisher information matrix [1, 2]. It is also a standard part of the information-theoretic framework surrounding Fisher information and entropy inequalities [3].

Let X be a random vector in $\mathbb{R}^n$ with density f, and write

$$\rho_f(x) = -\nabla \log f(x)$$

for its score function when the density is smooth and positive. The weak formulation used in the paper is stated in Section 2. The Fisher information matrix is

$$J(f) = \int_{\mathbb{R}^n} \rho_f(x)\rho_f(x)^{\mathsf{T}} f(x)\,dx.$$

If $\mu = \mathbb{E}X$, and

$$\Sigma = \mathbb{E}[(X-\mu)(X-\mu)^{\mathsf{T}}]$$

is the covariance matrix, then the matrix Cramér–Rao inequality states, when $J(f)$ is positive definite, that

$$\Sigma \geq J(f)^{-1}. \tag{1.1}$$

A standard scalar trace consequence is

$$\operatorname{tr}(\Sigma)^{1/2} \operatorname{tr}(J(f))^{1/2} \geq n. \tag{1.2}$$

The scalar inequality (1.2) is sharp, but it does not retain the full affine content of the matrix inequality (1.1). The matrix inequality reflects the intrinsic geometry of the pair $(\Sigma, J(f))$: Under a nonsingular linear change of position, the covariance matrix and the Fisher information matrix transform dually, and the full Gaussian equality class is preserved. The trace product in (1.2), however, is tied to the particular Euclidean position in which the random vector is written. Equality in the matrix inequality holds for every nonsingular Gaussian distribution, whereas equality in the scalar trace inequality holds only for those Gaussian distributions whose covariance is a scalar multiple of the identity in the chosen coordinates.

Thus the ordinary trace product is not the most natural scalar quantity when no preferred Euclidean position is specified. This point is closely related to affine information theory, where the aim is to formulate information quantities independent of arbitrary coordinates or weights.

Several distinct branches of affine Cramér–Rao theory should be distinguished. The star-body formulation of Lutwak, Yang, and Zhang [4] initiated a convex-geometric extension of the quadratic inequality. Their moment–entropy program subsequently treated Rényi entropy, generalized moments, and generalized Fisher information [5–7]. A second line introduced affine Fisher information by combining directional score information with affine isoperimetric quantities and obtained affine Fisher-information, Stam, and Cramér–Rao inequalities [8, 9]. Cianchi, Lutwak, Yang, and Zhang later placed a broad family of such inequalities in a unified convex-body framework [10]. General Fisher information matrices [11], recent information-geometric work on affine-deformation statistical models [12], and the extension of affine moment–entropy inequalities to matrix spaces [13] provide further developments.

Those constructions modify or enlarge the moment functional, the information functional, or both through directional or convex-geometric data. The present construction is complementary: It retains the classical quadratic second-moment matrix and the classical Fisher information matrix, regards them as a dual matrix pair, and removes the arbitrary Euclidean position by optimizing their trace product over $\mathrm{GL}(n)$. Thus no new affine Fisher information is introduced here.

More precisely, given positive definite matrices S and T, we minimize

$$\operatorname{tr}(ASA^{\mathsf{T}})^{1/2} \operatorname{tr}(A^{-\mathsf{T}}TA^{-1})^{1/2}, \qquad A \in \mathrm{GL}(n). \tag{1.3}$$

The first factor measures the quadratic size of the error after applying A, whereas the second measures the corresponding dual quadratic size of the score. Multiplying A by a positive scalar multiplies one factor and divides the other by the same scalar, so the product is scale invariant. Hence the minimization may equivalently be taken over volume-preserving linear maps.

The matrix ingredient is the following real positive-definite form of Alberti's variational characterization of matrix fidelity [14, 15].

Theorem 1.1. *Let S, T be positive definite symmetric $n \times n$ matrices. Then*

$$\inf_{A\in \mathrm{GL}(n)} \operatorname{tr}(ASA^{\mathsf{T}})^{1/2} \operatorname{tr}(A^{-\mathsf{T}}TA^{-1})^{1/2} = \operatorname{tr}\left[\left(S^{1/2}TS^{1/2}\right)^{1/2}\right]. \tag{1.4}$$

Moreover, if $P = A^{\mathsf{T}}A$, then A is optimal if and only if

$$P = cS^{-1/2}\left(S^{1/2}TS^{1/2}\right)^{1/2}S^{-1/2} \tag{1.5}$$

for some $c > 0$.

For a real matrix M, its trace norm is

$$\|M\|_1 = \operatorname{tr}|M|, \qquad |M| = (M^{\mathsf{T}}M)^{1/2};$$

equivalently, $\|M\|_1$ is the sum of the singular values of M. Taking $M = T^{1/2}S^{1/2}$ gives

$$|M| = (S^{1/2}TS^{1/2})^{1/2}, \qquad \|T^{1/2}S^{1/2}\|_1 = \operatorname{tr}\left[\left(S^{1/2}TS^{1/2}\right)^{1/2}\right].$$

The right-hand side of (1.4) is therefore the matrix fidelity of S and T, in the convention in which fidelity is not squared. It also appears in the Bures–Wasserstein geometry of positive definite matrices [16]; recent work on geometries of symmetric positive definite matrices gives additional context [17]. Alberti's formula states that

$$\left(\operatorname{tr}\left[\left(S^{1/2}TS^{1/2}\right)^{1/2}\right]\right)^2 = \inf_{P>0} \operatorname{tr}(SP)\operatorname{tr}(TP^{-1}). \tag{1.6}$$

Since every $P > 0$ can be written as $P = A^{\mathsf{T}}A$, formula (1.4) is precisely this classical identity in the real positive-definite setting. We do not claim a new variational characterization of the trace norm. The contribution pursued here is its Cramér–Rao interpretation, the description of the associated affine positions, and the resulting scalar equality structure.

We apply this matrix scalarization to random vectors. For $a \in \mathbb{R}^n$, set

$$\Sigma_a = \mathbb{E}[(X-a)(X-a)^{\mathsf{T}}], \qquad J(f) = \mathbb{E}[\rho_f(X)\rho_f(X)^{\mathsf{T}}],$$

and define the affine quadratic Cramér–Rao functional by

$$C_{\mathrm{aff}}(X,a) = \inf_{A\in \mathrm{GL}(n)} \left(\mathbb{E}|A(X-a)|^2\right)^{1/2}\left(\mathbb{E}|A^{-\mathsf{T}}\rho_f(X)|^2\right)^{1/2}. \tag{1.7}$$

Equivalently, because of scale invariance, the infimum may be taken over $A \in \mathrm{GL}(n)$ with $|\det A| = 1$.

The following theorem gives the probabilistic consequence of the matrix scalarization.

Theorem 1.2. *Let X be an $\mathbb{R}^n$-valued random vector with an admissible density f in the sense of Section 2, and assume that $J(f)$ is positive definite. Then, for every $a \in \mathbb{R}^n$,*

$$C_{\mathrm{aff}}(X, a) = \operatorname{tr}\left[\left(\Sigma_a^{1/2} J(f) \Sigma_a^{1/2}\right)^{1/2}\right]. \tag{1.8}$$

Moreover,

$$n \le C_{\mathrm{aff}}(X, a) \le (\operatorname{tr} \Sigma_a)^{1/2} (\operatorname{tr} J(f))^{1/2}. \tag{1.9}$$

Theorem 1.2 is proved in Section 4, where Theorem 4.1 also gives the complete equality characterization. The right-hand inequality in (1.9) only records that an infimum over affine positions is no larger than the value at $A = I_n$. Accordingly, C_{aff} is an optimized affine normalization of the fixed-coordinate trace product, not a logically stronger consequence of the matrix Cramér–Rao inequality. Its advantage is invariance under nonsingular affine changes of variables and a scalar equality condition that retains the complete nonsingular Gaussian family. Equality in the upper bound holds precisely when the original Euclidean position is already optimal, or equivalently when

$$\frac{\Sigma_a}{\operatorname{tr} \Sigma_a} = \frac{J(f)}{\operatorname{tr} J(f)}.$$

If f is a nonsingular Gaussian density and $a = \mathbb{E}X$, then $J(f) = \Sigma_a^{-1}$, so (1.8) gives $C_{\mathrm{aff}}(X, a) = n$, regardless of the covariance shape. By contrast, the usual trace inequality has equality only for isotropic Gaussians in the chosen Euclidean coordinates.

The paper is organized as follows. Section 2 demonstrates the score identity and the matrix Cramér–Rao inequality under Sobolev hypotheses. Section 3 gives a self-contained proof of the fidelity formula in the present notation and characterizes the optimal affine positions. Section 4 applies it to random vectors and discusses both equality cases.

2. Preliminaries

Throughout the paper, $n \ge 1$ is fixed. We write $\langle x, y \rangle$ for the Euclidean inner product in $\mathbb{R}^n$, $|x| = \langle x, x \rangle^{1/2}$, and $A^{-\mathsf{T}} = (A^{-1})^{\mathsf{T}}$. For symmetric matrices M, N, the notation $M \ge N$ means that $M - N$ is positive semidefinite.

For real matrices U, V, the Hilbert–Schmidt inner product and norm are

$$\langle U, V \rangle_{\mathrm{HS}} = \operatorname{tr}(U^{\mathsf{T}} V), \qquad \|U\|_{\mathrm{HS}}^2 = \operatorname{tr}(U^{\mathsf{T}} U). \tag{2.1}$$

A probability density f on $\mathbb{R}^n$ is called *admissible* if

$$f \ge 0, \qquad f \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n), \qquad \int_{\mathbb{R}^n} f(x)\, dx = 1,$$

and

$$\int_{\mathbb{R}^n} |x|^2 f(x)\, dx < \infty, \qquad \int_{\{f>0\}} \frac{|\nabla f(x)|^2}{f(x)}\, dx < \infty. \tag{2.2}$$

Define the score by

$$\rho_f(x) = \begin{cases} -\nabla f(x)/f(x), & f(x) > 0, \\ 0, & f(x) = 0. \end{cases} \tag{2.3}$$

For a Sobolev function, the weak gradient vanishes almost everywhere on each level set; in particular, $\nabla f = 0$ almost everywhere on $\{f = 0\}$. Hence $\rho_f f = -\nabla f$ almost everywhere. The Fisher information matrix is

$$J(f) = \mathbb{E}[\rho_f(X)\rho_f(X)^{\mathsf{T}}] = \int_{\mathbb{R}^n} \rho_f(x)\rho_f(x)^{\mathsf{T}} f(x)\,dx. \tag{2.4}$$

These hypotheses include the locally absolutely continuous formulation in one dimension and require neither global positivity nor a pointwise boundary-flux condition.

Lemma 2.1. *Let X have an admissible density f. Then, for every $a \in \mathbb{R}^n$,*

$$\mathbb{E}[(X - a)\rho_f(X)^{\mathsf{T}}] = I_n. \tag{2.5}$$

Proof. Choose $\eta \in C_c^\infty(\mathbb{R}^n)$ such that $0 \le \eta \le 1$, $\eta = 1$ on B_1, and $\eta = 0$ outside B_2, and put $\eta_R(x) = \eta(x/R)$. Then $|\nabla \eta_R| \le C/R$. Fix $1 \le i, j \le n$. Since $\rho_{f,j} f = -\partial_j f$ almost everywhere, distributional integration by parts gives

$$\begin{aligned}
&\int_{\mathbb{R}^n} \eta_R(x)(x_i - a_i)\rho_{f,j}(x) f(x)\,dx \\
=& -\int_{\mathbb{R}^n} \eta_R(x)(x_i - a_i)\partial_j f(x)\,dx \\
=& \delta_{ij} \int_{\mathbb{R}^n} \eta_R(x) f(x)\,dx + \int_{\mathbb{R}^n} (x_i - a_i)\partial_j \eta_R(x) f(x)\,dx.
\end{aligned} \tag{2.6}$$

The integrand on the left is absolutely integrable because

$$\int_{\mathbb{R}^n} |x_i - a_i|\,|\rho_{f,j}(x)| f(x)\,dx \le \left(\int_{\mathbb{R}^n} |x - a|^2 f(x)\,dx\right)^{1/2} \left(\int_{\mathbb{R}^n} |\rho_{f,j}(x)|^2 f(x)\,dx\right)^{1/2} < \infty.$$

Moreover, the cutoff error satisfies

$$\left|\int_{\mathbb{R}^n} (x_i - a_i)\partial_j \eta_R(x) f(x)\,dx\right| \le C\left(2 + \frac{|a|}{R}\right) \int_{\{R \le |x| \le 2R\}} f(x)\,dx,$$

which tends to zero. Letting $R \to \infty$ in (2.6) and using dominated convergence gives $\mathbb{E}[(X_i - a_i)\rho_{f,j}(X)] = \delta_{ij}$. This proves (2.5). □

Remark 2.2. *The admissibility conditions are convenient sufficient assumptions. The subsequent matrix and scalar inequalities use only finite second moments, finite Fisher information, and the integration-by-parts identity* (2.5)*. Thus they remain valid under any still weaker hypothesis that guarantees those three facts. The cutoff proof above shows explicitly why the earlier assumptions $f > 0$, $f \in C^1(\mathbb{R}^n)$, and a surface boundary-decay condition are not needed.*

For $a \in \mathbb{R}^n$, define the second-moment matrix about a by

$$\Sigma_a = \mathbb{E}[(X - a)(X - a)^{\mathsf{T}}]. \tag{2.7}$$

If $\mu = \mathbb{E}X$, we reserve the unsubscripted notation exclusively for the covariance matrix,

$$\Sigma = \Sigma_\mu = \mathrm{Cov}(X). \tag{2.8}$$

Throughout, we use S, T to denote abstract positive definite matrices; the notation $\Sigma_a, J(f)$ is used only in the probabilistic application.

Proposition 2.3. *Let X have an admissible density f, and assume that $J(f)$ is positive definite. Then*

$$\Sigma \geq J(f)^{-1}. \tag{2.9}$$

Consequently, for every $a \in \mathbb{R}^n$,

$$\Sigma_a \geq J(f)^{-1}. \tag{2.10}$$

Proof. By Lemma 2.1, with $a = \mu$,

$$\mathbb{E}[(X-\mu)\rho_f(X)^{\mathsf{T}}] = I_n.$$

Therefore the block second-moment matrix

$$\mathbb{E}\left[\begin{pmatrix} X-\mu \\ \rho_f(X) \end{pmatrix}\begin{pmatrix} X-\mu \\ \rho_f(X) \end{pmatrix}^{\mathsf{T}}\right] = \begin{pmatrix} \Sigma & I_n \\ I_n & J(f) \end{pmatrix}$$

is positive semidefinite. Since $J(f)$ is positive definite, its Schur complement gives $\Sigma - J(f)^{-1} \geq 0$. Finally,

$$\Sigma_a = \Sigma + (\mu - a)(\mu - a)^{\mathsf{T}} \geq \Sigma,$$

which proves (2.10). □

The usual scalar trace inequality follows from (2.9). Indeed, $J(f)^{1/2}\Sigma J(f)^{1/2} \geq I_n$. The matrices $J(f)^{1/2}\Sigma J(f)^{1/2}$ and $\Sigma^{1/2}J(f)\Sigma^{1/2}$ have the same eigenvalues, so

$$\operatorname{tr}\left[\left(\Sigma^{1/2}J(f)\Sigma^{1/2}\right)^{1/2}\right] \geq n.$$

On the other hand, the Schatten Cauchy–Schwarz inequality [18] yields

$$\operatorname{tr}\left[\left(\Sigma^{1/2}J(f)\Sigma^{1/2}\right)^{1/2}\right] = \|J(f)^{1/2}\Sigma^{1/2}\|_1 \leq \|J(f)^{1/2}\|_2\|\Sigma^{1/2}\|_2.$$

Consequently,

$$n \leq \operatorname{tr}\left[\left(\Sigma^{1/2}J(f)\Sigma^{1/2}\right)^{1/2}\right] \leq \operatorname{tr}(\Sigma)^{1/2}\operatorname{tr}(J(f))^{1/2}. \tag{2.11}$$

3. Affine matrix scalarization

This section recalls the matrix optimization result needed for the statistical application. Viewed solely as a matrix identity, it is the real positive-definite specialization of Alberti's variational formula (1.6); see [14, 15]. We include the short proof because it makes the full minimizer set and the balanced affine position explicit.

Let S, T be positive definite symmetric $n \times n$ matrices, and define

$$C(S,T) = \inf_{A \in \mathrm{GL}(n)} \operatorname{tr}(ASA^{\mathsf{T}})^{1/2}\operatorname{tr}(A^{-\mathsf{T}}TA^{-1})^{1/2}. \tag{3.1}$$

The product is unchanged when A is replaced by cA, $c > 0$. Thus the same infimum is obtained under the normalization $|\det A| = 1$.

Theorem 3.1. *Let S, T be positive definite symmetric $n \times n$ matrices. Then*

$$C(S,T) = \operatorname{tr}\left[\left(S^{1/2}TS^{1/2}\right)^{1/2}\right]. \tag{3.2}$$

Moreover, if $P = A^{\mathsf{T}}A$, then A is a minimizer in (3.1) *if and only if*

$$P = cS^{-1/2}\left(S^{1/2}TS^{1/2}\right)^{1/2}S^{-1/2} \tag{3.3}$$

for some $c > 0$. If $|\det A| = 1$, the condition $\det P = 1$ uniquely determines c.

Proof. Set $P = A^{\mathsf{T}}A$. Then P ranges over all positive definite symmetric matrices, and cyclicity of the trace gives

$$\operatorname{tr}(ASA^{\mathsf{T}}) = \operatorname{tr}(SP), \qquad \operatorname{tr}(A^{-\mathsf{T}}TA^{-1}) = \operatorname{tr}(TP^{-1}).$$

Consequently,

$$C(S,T)^2 = \inf_{P>0} \operatorname{tr}(SP)\operatorname{tr}(TP^{-1}). \tag{3.4}$$

Put

$$Q = S^{1/2}PS^{1/2}, \qquad K = S^{1/2}TS^{1/2}.$$

Equivalently,

$$P = S^{-1/2}QS^{-1/2}, \qquad P^{-1} = S^{1/2}Q^{-1}S^{1/2}.$$

Therefore, by cyclicity of the trace,

$$\operatorname{tr}(SP) = \operatorname{tr}(Q), \qquad \operatorname{tr}(TP^{-1}) = \operatorname{tr}\left(S^{1/2}TS^{1/2}Q^{-1}\right) = \operatorname{tr}(KQ^{-1}).$$

Then

$$C(S,T)^2 = \inf_{Q>0} \operatorname{tr}(Q)\operatorname{tr}(KQ^{-1}). \tag{3.5}$$

For every $Q > 0$, first write

$$\begin{aligned} \operatorname{tr}(K^{1/2}) &= \operatorname{tr}\left(Q^{1/2}Q^{-1/2}K^{1/2}\right) \\ &= \left\langle Q^{1/2}, Q^{-1/2}K^{1/2}\right\rangle_{\mathrm{HS}}. \end{aligned} \tag{3.6}$$

Applying Cauchy–Schwarz for the Hilbert–Schmidt inner product (2.1) gives

$$\begin{aligned} \operatorname{tr}(K^{1/2})^2 &\le \|Q^{1/2}\|_{\mathrm{HS}}^2 \|Q^{-1/2}K^{1/2}\|_{\mathrm{HS}}^2 \\ &= \operatorname{tr}(Q)\operatorname{tr}\left(K^{1/2}Q^{-1}K^{1/2}\right) \\ &= \operatorname{tr}(Q)\operatorname{tr}(KQ^{-1}). \end{aligned} \tag{3.7}$$

It follows from (3.5) that $C(S,T) \ge \operatorname{tr}(K^{1/2})$.

Equality in (3.7) holds if and only if the two matrices in (3.6) are positively linearly dependent, that is,

$$Q^{-1/2}K^{1/2} = \lambda Q^{1/2}$$

for some $\lambda > 0$. Equivalently, $Q = \lambda^{-1}K^{1/2}$. Hence the lower bound is attained and

$$C(S,T) = \operatorname{tr}(K^{1/2}) = \operatorname{tr}\left[\left(S^{1/2}TS^{1/2}\right)^{1/2}\right].$$

Since $Q = S^{1/2}PS^{1/2}$, the equality condition is precisely (3.3), with $c = \lambda^{-1}$. The determinant normalization uniquely fixes c. □

Remark 3.2. *Formula* (3.2)*, considered in isolation, is not claimed as new. Equation* (3.4) *identifies it exactly with Alberti's fidelity formula. The role of Theorem 3.1 here is to expose the minimizing congruence positions used in the Cramér–Rao application.*

Corollary 3.3. *Let S, T be positive definite symmetric matrices and $A \in \mathrm{GL}(n)$. Then A is a minimizer in* (3.1) *if and only if*

$$\frac{ASA^{\mathsf{T}}}{\mathrm{tr}(ASA^{\mathsf{T}})} = \frac{A^{-\mathsf{T}}TA^{-1}}{\mathrm{tr}(A^{-\mathsf{T}}TA^{-1})}. \tag{3.8}$$

Proof. Let $P = A^{\mathsf{T}}A$. By Theorem 3.1, A is optimal if and only if $PSP = c^2T$ for some $c > 0$. Multiplying on the left and right by $P^{-1/2}$ gives

$$P^{1/2}SP^{1/2} = c^2P^{-1/2}TP^{-1/2}.$$

Taking traces yields the scalar identity

$$\mathrm{tr}(P^{1/2}SP^{1/2}) = c^2\,\mathrm{tr}(P^{-1/2}TP^{-1/2}).$$

Dividing the preceding matrix identity by this trace identity gives

$$\frac{P^{1/2}SP^{1/2}}{\mathrm{tr}(P^{1/2}SP^{1/2})} = \frac{P^{-1/2}TP^{-1/2}}{\mathrm{tr}(P^{-1/2}TP^{-1/2})}.$$

Writing the polar decomposition $A = OP^{1/2}$, with O orthogonal, and conjugating both sides by O yields (3.8).

Conversely, (3.8), after removing the common orthogonal conjugation, implies

$$P^{1/2}SP^{1/2} = \gamma P^{-1/2}TP^{-1/2}, \qquad \gamma = \frac{\mathrm{tr}(P^{1/2}SP^{1/2})}{\mathrm{tr}(P^{-1/2}TP^{-1/2})} > 0,$$

and hence $PSP = \gamma T$. Multiplication by $S^{1/2}$ on both sides gives

$$(S^{1/2}PS^{1/2})^2 = \gamma S^{1/2}TS^{1/2}.$$

Uniqueness of the positive square root gives (3.3), with $c = \sqrt{\gamma}$, so A is a minimizer. □

Corollary 3.4. *Let S, T be positive definite symmetric matrices. Then*

$$C(S,T) \le \mathrm{tr}(S)^{1/2}\,\mathrm{tr}(T)^{1/2}. \tag{3.9}$$

Equality holds if and only if

$$\frac{S}{\mathrm{tr}\,S} = \frac{T}{\mathrm{tr}\,T}. \tag{3.10}$$

Proof. The inequality follows from (3.1) by taking $A = I_n$. Equality holds exactly when I_n is a minimizer, which, by Corollary 3.3, is equivalent to (3.10). □

4. Applications: The affine quadratic Cramér–Rao inequalities

Let X be an $\mathbb{R}^n$-valued random vector with admissible density f. For $a \in \mathbb{R}^n$, set

$$\Sigma_a = \mathbb{E}[(X-a)(X-a)^{\mathsf{T}}], \qquad J(f) = \mathbb{E}[\rho_f(X)\rho_f(X)^{\mathsf{T}}],$$

and define

$$C_{\mathrm{aff}}(X,a) = \inf_{A\in \mathrm{GL}(n)} (\mathbb{E}|A(X-a)|^2)^{1/2} (\mathbb{E}|A^{-\mathsf{T}}\rho_f(X)|^2)^{1/2}. \tag{4.1}$$

This functional is invariant under nonsingular affine changes of variables. Indeed, let $Y = BX + b$, where $B \in \mathrm{GL}(n)$, and let f_Y denote the density of Y. Then

$$\Sigma_{Y,b+Ba} = B\Sigma_{X,a}B^{\mathsf{T}}, \qquad \rho_{f_Y}(Y) = B^{-\mathsf{T}}\rho_f(X), \qquad J(f_Y) = B^{-\mathsf{T}}J(f)B^{-1}.$$

Therefore, with $G = AB$,

$$\begin{aligned} C_{\mathrm{aff}}(Y, b+Ba) &= \inf_{A\in \mathrm{GL}(n)} (\mathbb{E}|AB(X-a)|^2)^{1/2} (\mathbb{E}|A^{-\mathsf{T}}B^{-\mathsf{T}}\rho_f(X)|^2)^{1/2} \\ &= \inf_{G\in \mathrm{GL}(n)} (\mathbb{E}|G(X-a)|^2)^{1/2} (\mathbb{E}|G^{-\mathsf{T}}\rho_f(X)|^2)^{1/2} \\ &= C_{\mathrm{aff}}(X,a). \end{aligned}$$

Scale invariance in A also shows that the infimum may be restricted to $|\det A| = 1$.

Theorem 4.1. *Let X have an admissible density f on $\mathbb{R}^n$, let $\mu = \mathbb{E}X$, and assume that $J(f)$ is positive definite. Then, for every $a \in \mathbb{R}^n$,*

$$C_{\mathrm{aff}}(X,a) = \mathrm{tr}\left[\left(\Sigma_a^{1/2} J(f) \Sigma_a^{1/2}\right)^{1/2}\right]. \tag{4.2}$$

Moreover,

$$n \le C_{\mathrm{aff}}(X,a) \le \mathrm{tr}(\Sigma_a)^{1/2}\, \mathrm{tr}(J(f))^{1/2}. \tag{4.3}$$

Equality in the lower bound holds if and only if $a = \mu$ and f agrees almost everywhere with a nonsingular Gaussian density centered at a. Equivalently,

$$f(x) = C\exp\{-\lambda|B(x-a)|^2\}, \qquad \lambda > 0, \quad B \in \mathrm{GL}(n), \tag{4.4}$$

where $C = |\det B|(\lambda/\pi)^{n/2}$. Equality in the upper bound holds if and only if

$$\frac{\Sigma_a}{\mathrm{tr}\,\Sigma_a} = \frac{J(f)}{\mathrm{tr}\,J(f)}. \tag{4.5}$$

Proof. For every $A \in \mathrm{GL}(n)$,

$$\mathbb{E}|A(X-a)|^2 = \mathrm{tr}(A\Sigma_a A^{\mathsf{T}}), \qquad \mathbb{E}|A^{-\mathsf{T}}\rho_f(X)|^2 = \mathrm{tr}(A^{-\mathsf{T}}J(f)A^{-1}).$$

Thus (4.2) is Theorem 3.1 with $S = \Sigma_a$ and $T = J(f)$.

By Proposition 2.3, $\Sigma_a \ge J(f)^{-1}$, or equivalently

$$J(f)^{1/2}\Sigma_a J(f)^{1/2} \ge I_n.$$

The matrices $J(f)^{1/2}\Sigma_a J(f)^{1/2}$ and $\Sigma_a^{1/2} J(f)\Sigma_a^{1/2}$ have the same positive eigenvalues. Hence every eigenvalue of the latter matrix is at least one, and (4.2) yields the lower bound. The upper bound and its equality condition follow from Corollary 3.4.

It remains to characterize equality in the lower bound. Suppose $C_{\mathrm{aff}}(X, a) = n$. The preceding eigenvalue argument shows that

$$\Sigma_a^{1/2} J(f)\Sigma_a^{1/2} = I_n,$$

and hence $\Sigma_a = J(f)^{-1}$. Since

$$\Sigma_a = \Sigma + (\mu - a)(\mu - a)^{\mathsf{T}}, \qquad \Sigma \succeq J(f)^{-1},$$

we obtain

$$(\Sigma - J(f)^{-1}) + (\mu - a)(\mu - a)^{\mathsf{T}} = 0.$$

Both summands are positive semidefinite, so

$$\Sigma = J(f)^{-1}, \qquad a = \mu. \tag{4.6}$$

By the score identity, $\mathbb{E}[(X - \mu)\rho_f(X)^{\mathsf{T}}] = I_n$. Set

$$Z = X - \mu - J(f)^{-1}\rho_f(X).$$

A direct expansion using (4.6) gives $\mathbb{E}[ZZ^{\mathsf{T}}] = 0$. Taking traces yields $\mathbb{E}|Z|^2 = 0$, so

$$\rho_f(X) = J(f)(X - \mu) \qquad \text{almost surely.} \tag{4.7}$$

We now justify Gaussianity without invoking an unproved passage from almost-sure to pointwise equality. Equation (4.7) means $\rho_f(x) = J(f)(x-\mu)$ for $f(x)\,dx$-almost every x. Multiplying by f, using $\rho_f f = -\nabla f$ on $\{f > 0\}$, and using $\nabla f = 0$ almost everywhere on $\{f = 0\}$, we obtain the weak-gradient identity

$$\nabla f(x) = -J(f)(x - \mu)f(x) \qquad \text{for Lebesgue-almost every } x \in \mathbb{R}^n. \tag{4.8}$$

Define

$$h(x) = \exp\left\{\frac{1}{2}(x - \mu)^{\mathsf{T}} J(f)(x - \mu)\right\} f(x).$$

Then $h \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$, and the Sobolev product rule together with (4.8) gives $\nabla h = 0$ weakly. A locally Sobolev function with zero weak gradient is constant almost everywhere on the connected set $\mathbb{R}^n$. Therefore

$$f(x) = C \exp\left\{-\frac{1}{2}(x - \mu)^{\mathsf{T}} J(f)(x - \mu)\right\} \quad \text{a.e.,} \tag{4.9}$$

where normalization gives $C = (2\pi)^{-n/2} \det(J(f))^{1/2}$. Thus f is a nonsingular Gaussian density and $a = \mu$.

Conversely, if (4.4) holds, then $a = \mathbb{E}X$ and $J(f) = \Sigma_a^{-1}$. Equation (4.2) gives $C_{\mathrm{aff}}(X, a) = n$. □

Example 4.2 (A two-dimensional anisotropic Gaussian)**.** *Let X be a centered Gaussian vector in $\mathbb{R}^2$ with covariance matrix*

$$\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$J(f) = \Sigma^{-1} = \begin{pmatrix} 1/4 & 0 \\ 0 & 1 \end{pmatrix},$$

and the fixed-coordinate trace product is

$$\operatorname{tr}(\Sigma)^{1/2} \operatorname{tr}(J(f))^{1/2} = \sqrt{5}\sqrt{\frac{5}{4}} = \frac{5}{2}.$$

Now take the volume-preserving matrix

$$A = \begin{pmatrix} 2^{-1/2} & 0 \\ 0 & 2^{1/2} \end{pmatrix}, \qquad \det A = 1.$$

A direct calculation gives

$$A\Sigma A^{\mathsf{T}} = 2I_2, \qquad A^{-\mathsf{T}} J(f) A^{-1} = \frac{1}{2} I_2.$$

Consequently,

$$\left[\operatorname{tr}(A\Sigma A^{\mathsf{T}})\right]^{1/2} \left[\operatorname{tr}(A^{-\mathsf{T}} J(f) A^{-1})\right]^{1/2} = \sqrt{4}\sqrt{1} = 2.$$

The lower bound in Theorem 4.1 shows that this value is optimal. Hence

$$C_{\mathrm{aff}}(X, 0) = 2,$$

whereas the trace product in the original anisotropic coordinates is $5/2$. *Thus the affine minimization recovers the Gaussian equality value even though* Σ *is not a scalar multiple of the identity.*

The inequality (4.3) should be interpreted as an optimized affine reformulation of the same matrix Cramér–Rao information. The relation $C_{\mathrm{aff}}(X, a) \le \operatorname{tr}(\Sigma_a)^{1/2} \operatorname{tr}(J(f))^{1/2}$ follows because the identity position is one competitor in the infimum; it is not a claim of greater logical strength. The normalization advantage appears in the equality cases. For a Gaussian centered at a, the fixed-coordinate trace product equals

$$\operatorname{tr}(\Sigma_a)^{1/2} \operatorname{tr}(\Sigma_a^{-1})^{1/2},$$

which is n only if Σ_a is a scalar multiple of I_n, whereas $C_{\mathrm{aff}} = n$ for every nonsingular covariance matrix. Condition (4.5) says that the normalized covariance and Fisher-information shapes already coincide in the original Euclidean coordinates. Equivalently, I_n is already a minimizing affine position. Hence no nontrivial anisotropic affine transformation is needed to attain the minimum; under the normalization $|\det A| = 1$, every orthogonal transformation is also minimizing.

5. Conclusions

The matrix Cramér–Rao inequality is compatible with nonsingular affine changes of variables, but its usual scalar trace consequence depends on the Euclidean coordinates in which it is stated. In this paper, we removed this dependence by minimizing the trace product over all nonsingular linear positions. The classical variational formula for matrix fidelity gives the exact value of this minimum and also allows all minimizing positions to be determined. Since the construction uses the usual

covariance matrix and the usual Fisher information matrix, it does not require a modified notion of Fisher information.

For an admissible random vector, the resulting quantity is affine invariant and satisfies

$$n \le C_{\mathrm{aff}}(X, a) \le \mathrm{tr}(\Sigma_a)^{1/2}\,\mathrm{tr}(J(f))^{1/2}.$$

The upper bound simply compares the optimal affine position with the original one. In particular, it identifies when the given Euclidean coordinates are already optimal. More importantly, equality in the lower bound holds exactly when a is the mean and the distribution is a nonsingular Gaussian. Thus every nonsingular Gaussian, including one with an anisotropic covariance matrix, has the same sharp value n. The affine minimization therefore restores, at the scalar level, the full Gaussian equality class of the matrix Cramér–Rao inequality, which is not retained by the fixed-coordinate trace product.

Use of Generative-AI tools declaration

The author declares that DeepSeek-V4 was used only for language polishing and editorial refinement of this manuscript. The author reviewed and revised all AI-assisted text and takes full responsibility for the content of the article.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grant No. 12571146.

Conflict of interest

The author declares that there is no conflict of interest.

References

1. H. Cramér, *Mathematical methods of statistics*, Princeton: Princeton University Press, 1999.
2. C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, In: *Breakthroughs in statistics*, New York: Springer, 1992. https://doi.org/10.1007/978-1-4612-0919-5_16
3. T. M. Cover, J. A. Thomas, *Elements of information theory*,New York: Wiley, 2005.
4. E. Lutwak, D. Yang, G. Zhang, The Cramér–Rao inequality for star bodies, *Duke Math. J.*, **112** (2002), 59–81. https://doi.org/10.1215/S0012-9074-02-11212-5
5. E. Lutwak, D. Yang, G. Zhang, Moment-entropy inequalities, *Ann. Probab.* **32** (2004), 757–774. https://doi.org/10.1214/aop/1079021463
6. E. Lutwak, D. Yang, G. Zhang, Cramér–Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information, *IEEE Trans. Inform. Theory*, **51** (2005), 473–478. https://doi.org/10.1109/TIT.2004.840871

7. E. Lutwak, D. Yang, G. Zhang, Moment-entropy inequalities for a random vector, *IEEE Trans. Inform. Theory*, **53** (2007), 1603–1607. https://doi.org/10.1109/TIT.2007.892780
8. E. Lutwak, S. Lv, D. Yang, G. Zhang, Extensions of Fisher information and Stam's inequality, *IEEE Trans. Inform. Theory*, **58** (2012), 1319–1327. https://doi.org/10.1109/TIT.2011.2177563
9. S. J. Lv, X. Lv, Affine Fisher information inequalities, *J. Math. Anal. Appl.*, **371** (2010), 347–354. https://doi.org/10.1016/j.jmaa.2010.05.034
10. A. Cianchi, E. Lutwak, D. Yang, G. Zhang, A unified approach to Cramér–Rao inequalities, *IEEE Trans. Inform. Theory*, **60** (2014), 643–650. https://doi.org/10.1109/TIT.2013.2284498
11. S. J. Lv, General Fisher information matrices of a random vector, *Adv. Appl. Math.*, **89** (2017), 18–40. https://doi.org/10.1016/j.aam.2017.03.002
12. S. Amari, T. Matsuda, Information geometry of Wasserstein statistics on shapes and affine deformations, *Inf. Geom.*, **7** (2024), 285–309. https://doi.org/10.1007/s41884-024-00139-y
13. D. Langharst, On moment–entropy inequalities in the space of matrices, *Electron. J. Probab.*, **31** (2026), 1–23. https://doi.org/10.1214/26-EJP1552
14. P. M. Alberti, A note on the transition probability over C^*-algebras, *Lett. Math. Phys.*, **7** (1983), 25–32. https://doi.org/10.1007/BF00398708
15. J. Watrous, *The theory of quantum information*, Cambridge: Cambridge University Press, 2018. https://doi.org/10.1017/9781316848142
16. R. Bhatia, T. Jain, Y. Lim, On the Bures-Wasserstein distance between positive definite matrices, *Expo. Math.*, **37** (2019), 165–191. https://doi.org/10.1016/j.exmath.2018.01.002
17. Y. Thanwerdas, X. Pennec, The geometry of mixed-Euclidean metrics on symmetric positive definite matrices, *Differential Geom. Appl.*, **81** (2022), 101867. https://doi.org/10.1016/j.difgeo.2022.101867
18. R. Bhatia, *Matrix analysis*, New York: Springer, 1997. https://doi.org/10.1007/978-1-4612-0653-8

AIMS Press