Quantifying Differences between High-Dimensional Beam Phase Space Distributions with f-Divergences

Authors: Du, Mr. Yu, Dr. Zhi-Jun Wang, Wong, Mr. Jonathan C., Tao Zhang, Ma, Ms. Bing-Hui, Zhou, Hao-Yu, Chun-Guang Su, Liu, Mr. Li-Wen, Ling-Yun Gong, Yi, Ms. Man, Chen, Dr. Xiaolong, Chi Feng, Dr. Zhi-Jun Wang

Date: 2025-09-13T23:00:32+00:00

Abstract

Quantifying differences between high-dimensional phase space distributions is essential for analyzing beam measurements and simulations. While f-divergences such as KL or JS divergence are increasingly used for this purpose, including in machine learning applications, their values lack physical interpretability. This work establishes the first physics-grounded framework for f-divergences in accelerator beam contexts. Through systematic analysis of 4D transverse phase space distributions with elliptical symmetry, we reveal how distinct f-divergences assign region-specific weights to distribution cores, tails, and halos. We also prove rigorous correspondences between f-divergence values and conventional beam physics quantities: emittance differences and mismatch factors. These results, validated by statistical analysis of synthetic distributions, provide concrete selection guidelines for f-divergences in phase space comparisons and establish assessment standards for evaluating f-divergence values in beam physics applications.

Full Text

Preamble

Quantifying Differences between High-Dimensional Beam Phase Space Distributions with f-Divergences[*†]

Jonathan C. Wong,[1,2,3,‡] Tao Zhang,[1,2,3] Bing-Hui Ma,[1,2,3] Hao-Yu Zhou,[1,2,3] Yu Du,[1,2,3] Zhi-Jun Wang,[1,2,3] Chun-Guang Su,[1,2,3] Li-Wen Liu,[1,2,3] Ling-Yun Gong,[1,2,3] Man Yi,[3] Xiao-Long Chen,[1,2,3] and Chi Feng[1,2,3]

[1]Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Advanced Energy Science and Technology Guangdong Laboratory, Huizhou 516000, China

Quantifying differences between high-dimensional phase space distributions is essential for analyzing beam measurements and simulations. While f-divergences such as KL or JS divergence are increasingly used for this purpose, including in machine learning applications, their values lack physical interpretability. This work establishes the first physics-grounded framework for f-divergences in accelerator beam contexts. Through systematic analysis of 4D transverse phase space distributions with elliptical symmetry, we reveal how distinct f-divergences assign region-specific weights to distribution cores, tails, and halos. We also prove rigorous correspondences between f-divergence values and conventional beam physics quantities: mismatch factors and emittance differences. These results, validated by statistical analysis of synthetic distributions, provide concrete selection guidelines for f-divergences in phase space comparisons and establish assessment standards for evaluating f-divergence values in beam physics applications.

Keywords: High-Dimensional Phase Space, f-Divergences, Mismatch Factor, Emittance Measurement

Introduction

In high-power accelerator beam experiments and simulations, accurately quantifying beam distribution differences in high-dimensional phase space is crucial for understanding beam transmission stability, optimizing beam matching strategies, and diagnosing beam anomalies. These are all key factors in ensuring beam quality and transmission efficiency.

The most commonly used method for quantifying beam distribution differences is currently the 2D mismatch factor, particularly in beamline design [**?**]. This method quantifies differences by comparing distribution contours in phase space, which are described by Twiss parameters and emittance. All such information is fully contained within the beam matrix, whose measurement techniques have been comprehensively explained in Ref. [**?**] on emittance measurement. However, the mismatch factor cannot distinguish between different types of beam distributions with identical Twiss parameters in 2D phase space, nor can it directly measure beam distribution differences in high-dimensional phase space.

To overcome these limitations, statistical divergences—widely used in statistics, information theory, and machine learning—can be adopted as quantitative tools for beam distribution differences. They comprehensively reflect the total contribution of differences at all points in phase space. Ref. [**?**] introduced the Kernel-Based Statistical Distance into beam physics research. This paper focuses on investigating f-divergences, primarily including the Kullback-Leibler (KL) divergence [**?**, **?**], Jensen-Shannon (JS) divergence [**?**, **?**], Total Variation (TV) distance [**?**, **?**], and Hellinger distance [**?**, **?**], all of which offer reasonable computational complexity. f-divergences provide additional tools for analyzing experimental or simulation data related to beam distributions. For example, in beam phase space tomography studies, f-divergences such as KL or JS divergence are often used to directly compare differences between reconstructed and true distributions in high-dimensional phase space, thereby evaluating reconstruction algorithm reliability and result correctness [?]. Nevertheless, f-divergences are currently mostly used as loss functions. It remains unclear how their applicability varies across different distribution regions, and their values lack intuitive physical interpretation.

This study aims to clarify the regional application emphasis of different f-divergences, reveal the physical meanings of their values in beam physics, and provide guidance for their effective application in this field.

The paper is organized as follows. Section II introduces several ideal beam distributions commonly used in accelerator beam simulation studies, along with their quadratic form expressions. Subsequently, it presents f-divergences for quantifying differences in high-dimensional phase space beam distributions and computational formulas for four specific types. The subsequent content develops discussions based on these distributions and divergences. For distributions of different types sharing identical RMS moments, Section III investigates the dependence of f-divergence values on RMS moments. Leveraging this relationship, functional curves of f-divergence versus integration radius are derived by simplifying complex distribution forms; Appendix A provides analytical expressions for some obtainable curves. Furthermore, f-divergence values satisfying this condition are computed in 4D transverse phase space. Section IV explores the regional applicability of different f-divergences for quantifying beam distribution differences in high-dimensional phase space. The phase space is divided into three regions based on equidensity points and boundary points of beam distributions: the beam core, tail, and halo. Subsequently, leveraging the f-divergence-integration radius curves obtained in Section III, the weight allocation of each divergence across these three regions is quantified, thereby revealing their respective regional applicability emphases.

For beam distributions with different RMS moments and no coupling in real space directions, the study comprises two parts. Section V proves in 2D phase space that a correspondence exists between f-divergence and mismatch factor for distributions sharing identical RMS emittance. This relationship is then generalized to 2n-D phase space, and its validity is confirmed via simulations in 4D transverse phase space. Section VI proves in 2D phase space that a correspondence exists between f-divergence of matched beam distributions and the RMS emittance scaling factor, and also generalizes this relationship to 2n-D phase space. Furthermore, simulations in 4D transverse phase space reveal a unique symmetry exhibited by the f-divergence under these conditions. Detailed proofs of certain results in Sections V and VI are included in Appendix B. Section VII establishes assessment standards for two f-divergences in 4D transverse phase space, based on the correspondences proven in Sections V and VI, thereby providing a basis for tracing the sources of beam differences.

II. Basic Terms

A. Common Beam Distributions

In accelerator beam simulation studies, several commonly used beam phase space distributions include the K-V, waterbag, parabolic, and Gaussian distributions, all of which exhibit elliptical symmetry [?]. For a 2n-D beam distribution satisfying this symmetry and centered at the origin of phase space coordinates, its normalized density function (we will omit the phrase "normalized" for brevity and take all distributions to be normalized in the rest of the paper) can be uniformly described by the following quadratic form expression:

$$\rho(x) = kg(x^T \Sigma^{-1} x)$$

where $x = (x_1, x_1', \dots, x_n, x_n')^T$ represents the 2n-D phase space coordinates of the beam. The 2n-D symmetric matrix:

$$\Sigma = \begin{pmatrix} \langle x_1 x_1 \rangle & \langle x_1 x_1' \rangle & \dots \\ \langle x_1' x_1 \rangle & \langle x_1' x_1' \rangle & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

is the covariance matrix of the beam distribution. It is positive definite and consists of $2n^2 + n$ independent second-order moments. If all second-order moments in its off-diagonal 2D submatrices are zero, then the beam distribution is said to be uncoupled in the motion directions. By introducing the covariance matrix, the Mahalanobis distance can be defined as:

$$d = \sqrt{x^T \Sigma^{-1} x}$$

This is a dimensionless quantity. At points with identical Mahalanobis distance values, the density of the beam distribution remains identical. Geometrically, when d is held constant, Eq. (3) represents a 2n-D ellipsoid centered at the origin of the phase space coordinates. The RMS emittance ε_{rms} of the beam distribution corresponds to the total emittance of the phase ellipsoid at $d = 1$, with its value given by $\varepsilon_{rms} = \sqrt{\det(\Sigma)}$. Finally, in Eq. (1), different coefficients k and functions $g(x^T \Sigma^{-1} x)$ correspond to different types of beam distributions, as detailed in Table 1.

As can be seen from Table 1, the K-V distribution is characterized by particles uniformly distributed on the surface of the 2n-D phase ellipsoid. This distribution property makes it difficult for f-divergence-based difference quantification methods to be effectively applied. Therefore, this study only considers the waterbag, parabolic, and Gaussian distributions, excluding the K-V distribution.

B. f-Divergences

The f-divergences are a class of methods for quantifying differences between two distributions [**?**], defined as:

$$D_f[\rho_1(x)\|\rho_2(x)] = \int \rho_2(x) f\left(\frac{\rho_1(x)}{\rho_2(x)}\right) dx$$

where $\rho_1(x)$ and $\rho_2(x)$ are the density functions of the two distributions; f is a convex function satisfying $f(1) = 0$, which ensures the non-negativity of D_f, i.e., $D_f \geq 0$, with $D_f = f(1) = 0$ when $\rho_1(x) = \rho_2(x)$. Different choices of f correspond to different statistical divergences.

This study employs four commonly used divergences for beam distribution analysis. The corresponding forms of f, the computational formulas for D_f, and their value ranges are listed in Table 2.

It should be noted that the KL divergence is asymmetric: $D_{KL}[\rho_1(x)\|\rho_2(x)] \neq D_{KL}[\rho_2(x)\|\rho_1(x)]$ and it violates the triangle inequality. Additionally, to ensure the KL divergence is well-defined, the conditions $\rho_1(x) = 0 \Rightarrow \rho_2(x) = 0$ must be satisfied. Therefore, the following three cases that violate these conditions require special treatment:

1. If $\rho_1(x) = \rho_2(x) = 0$, there is no difference at that point, and we define $h = 0$;
2. If $\rho_1(x) = 0$ and $\rho_2(x) \neq 0$, the limit value can be attained at that point: $h = \lim_{\rho_1(x)\to 0} \rho_1(x) \ln[\rho_1(x)] - \ln[\rho_2(x)] = 0$;
3. If $\rho_1(x) \neq 0$ and $\rho_2(x) = 0$, then at that point: $h = \lim_{\rho_2(x)\to 0} \rho_2(x) \ln\left(\frac{\rho_1(x)}{\rho_2(x)}\right)$.

Cases 2 and 3 correspond to regions where the beam distributions do not overlap. The value of the function h indicates that in such regions, the D_{KL} under Case 2 lacks statistical significance, while it is undefined under Case 3. In this paper, to ensure D_{KL} yields a well-defined value, we will either select the distribution with the broader range as ρ_2 for computing D_{KL} to satisfy the conditions of Case 2, or set ρ_2 to a very small non-zero value to avoid the occurrence of Case 3.

III. f-Divergences Between Different Types of Beam Distributions with Identical Covariance Matrices

A. Properties of f-Divergence Under Identical Σ Condition

As shown in Eq. (1), for particle beam distributions with elliptical symmetry, the constant k and the function $g(x^T\Sigma^{-1}x)$ determine the distribution type, while the covariance matrix Σ determines the corresponding geometric shape. The difference between two distributions thus stems from both their type and geometric shape. For beam distributions of different types but with identical geometric shapes, the f-divergences measuring their difference satisfy the following theorem:

Theorem 1. In the 2n-D phase space $(x_1, x_1', \ldots, x_n, x_n')$, for any two elliptically symmetric beam distributions of different types sharing the same covariance matrix Σ, their f-divergence is independent of the specific values of the second-order moments in Σ.

We now prove Theorem 1.

In the 2n-D phase space $(x_1, x_1', \ldots, x_n, x_n')$, consider any two distinct types of beam distributions $\rho_1(x)$ and $\rho_2(x)$ listed in Table 1 that share the same covariance matrix Σ. According to Eqs. (1) and (4), their f-divergences can be expressed as:

$$D_f[\rho_1(x)\|\rho_2(x)] = \int \rho_2(x) f\left(\frac{k_1 \cdot g_1(x^T\Sigma^{-1}x)}{k_2 \cdot g_2(x^T\Sigma^{-1}x)}\right) dx_1 dx_1' \ldots dx_n dx_n'$$

From Lemma 1 in Appendix B, the quadratic form beam distributions in the $(y_1, y_1', \ldots, y_n, y_n')$ phase space can be transformed into canonical form via the orthogonal transformation $x = Py$, where P diagonalizes Σ such that $P^T\Sigma P = \Lambda$. The relationship between volume elements before and after coordinate transformation is:

$$dx_1 dx_1' \ldots dx_n dx_n' = dy_1 dy_1' \ldots dy_n dy_n' \cdot |J|$$

where $|J|$ denotes the Jacobian determinant of the phase space coordinate transformation. Since this transformation is orthogonal, its value is $|J| = 1$.

Consequently, in the $(y_1, y_1', \ldots, y_n, y_n')$ phase space, Eq. (6) equivalently becomes:

$$D_f[\rho_1(Py)\|\rho_2(Py)] = \int k_2 \cdot g_2(y^T\Lambda^{-1}y) f\left(\frac{k_1 \cdot g_1(y^T\Lambda^{-1}y)}{k_2 \cdot g_2(y^T\Lambda^{-1}y)}\right) dy_1 dy_1' \ldots dy_n dy_n'$$

Next, we perform a normalization transformation on the $(y_1, y_1', \ldots, y_n, y_n')$ phase space:

$$z = Q^{-1}y$$

where $Q = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_1'}, \ldots, \sqrt{\lambda_n}, \sqrt{\lambda_n'})$ and λ_i are the eigenvalues of Σ. The relationship between volume elements before and after this coordinate transformation is:

$$dy_1 dy_1' \ldots dy_n dy_n' = dz_1 dz_1' \ldots dz_n dz_n' \cdot |J'|$$

where $|J'|$ is the Jacobian determinant of this phase space coordinate transformation. Since this is a linear transformation, its value is $|J'| = \det(Q) = \sqrt{\det(\Sigma)}$.

In the $(z_1, z_1', \dots, z_n, z_n')$ phase space, according to Eqs. (12) and (13), the quadratic form $d^2 = y^T \Lambda^{-1} y$ becomes:

$$d^2 = (Qz)^T \Lambda^{-1} (Qz) = z^T I^{-1} z = z^T z$$

where I denotes the 2n-D identity matrix. When $d = 1$, Eq. (16) geometrically represents a 2n-D RMS unit hypersphere centered at the origin of the phase space coordinates. Substituting Eqs. (12), (14), (15), and (16) into Eq. (11) yields:

$$D_f[\rho_1(PQz) \| \rho_2(PQz)] = \int k_2 \cdot g_2(z^T z) f\left(\frac{k_1 \cdot g_1(z^T z)}{k_2 \cdot g_2(z^T z)}\right) dz_1 dz_1' \dots dz_n dz_n'$$

From Eq. (17), it can be observed that after the two-step transformation defined by Eqs. (7) and (12), the f-divergence derived from Eq. (6) is equivalent to that computed directly in the $(z_1, z_1', \dots, z_n, z_n')$ phase space using Eq. (4). The density function of the beam distribution in the $(z_1, z_1', \dots, z_n, z_n')$ phase space is given by:

$$\rho(z) = k \cdot g(z^T z)$$

where the expressions for k and g refer to Table 1. The identity matrix I serves as the covariance matrix of the beam distribution, with its second-order moments being fixed. However, in the original $(x_1, x_1', \dots, x_n, x_n')$ phase space, the values of the second-order moments in the covariance matrix Σ are arbitrary. This demonstrates that for any two distinct types of beam distributions sharing the same covariance matrix Σ in the $(x_1, x_1', \dots, x_n, x_n')$ phase space, their f-divergence is independent of the specific values of the second-order moments in Σ. Q.E.D.

B. f-Divergence Values Between Different 4D Distributions Sharing Identical Σ

The conclusion at the end of Section III.A demonstrates that for the f-divergences between any different types of distributions sharing the same covariance matrix Σ, we need only solve the problem once by replacing Σ with the identity matrix I, after which these results can be directly applied.

To facilitate f-divergence computation, we further transform the $(z_1, z_1', \dots, z_n, z_n')$ phase space into spherical coordinates:

$$\begin{cases} z_1 = r\cos\psi_1 \\ z_1' = r\sin\psi_1\cos\psi_2 \\ \vdots \\ z_n = r\sin\psi_1 \dots \sin\psi_{2n-2}\cos\psi_{2n-1} \\ z_n' = r\sin\psi_1 \dots \sin\psi_{2n-2}\sin\psi_{2n-1} \end{cases}$$

with angular ranges: $\psi_1, \dots, \psi_{2n-2} \in [0, \pi]$ and $\psi_{2n-1} \in [0, 2\pi]$.

After this transformation, Eq. (16) becomes: $d^2 = z^T z = r^2$.

The volume element transforms as: $dz_1 \dots dz_n' = dr d\psi_1 \dots d\psi_{2n-1} \cdot |J_{sph}|$, where the Jacobian determinant is:

$$|J_{sph}| = r^{2n-1} \prod_{k=1}^{2n-2} \sin^{2n-1-k}(\psi_k)$$

Thus, in the $(r, \psi_1, \dots, \psi_{2n-1})$ coordinates, Eq. (17) becomes:

$$D_f[\rho_1(r) \| \rho_2(r)] = \int \rho_2(r) f\left(\frac{\rho_1(r)}{\rho_2(r)}\right) \prod_{k=1}^{2n-1} \sin^{2n-1-k}(\psi_k) d\psi_k dr$$

with the density function: $\rho(r) = k \cdot g(r^2)$.

For $n = 2$ (4D phase space), the specific forms of three distributions in $(r, \psi_1, \psi_2, \psi_3)$ coordinates are:

- **Waterbag:**

$$\rho_{(4D)}(r) = \begin{cases} \frac{2}{(\sqrt{6}\pi)^4}, & r^2 < 6 \\ 0, & \text{otherwise} \end{cases}$$

- **Parabolic:**

$$\rho_{(4D)}(r) = \begin{cases} \frac{6}{(\sqrt{8}\pi)^4} \cdot \left(1 - \frac{r^2}{8}\right), & r^2 < 8 \\ 0, & \text{otherwise} \end{cases}$$

- **Gaussian:**

$$\rho_{(4D)}(r) = \frac{1}{(\sqrt{2}\pi)^4} \cdot \exp\left(-\frac{r^2}{2}\right)$$

Substituting Eqs. (26), (27), and (28) into Eq. (24), we can compute the f-divergence values between different types of beam distributions in the $(r, \psi_1, \psi_2, \psi_3)$ phase space through numerical integration. This corresponds to the f-divergences between different types of beam distributions with identical covariance matrix Σ in the 4D transverse phase space (x_1, x_1', x_2, x_2'). The specific results are shown in Table 3.

The numerical integration is performed using a discrete approximation:

$$D_f[\rho_1(x) \| \rho_2(x)] \approx \sum_{i=1}^{m} \rho_2(x_i) f\left(\frac{\rho_1(x_i)}{\rho_2(x_i)}\right) \Delta V$$

where $x_i = (x_{1i}, x_{1i}', x_{2i}, x_{2i}')$ denotes the coordinates at the i-th node. The total number of nodes $m = (2l/d)^4$ determines the computational accuracy, with l and d representing the definition range along each coordinate axis and the discrete cell length, respectively. The discrete phase-space volume element $\Delta V = \Delta x_1 \Delta x_1' \Delta x_2 \Delta x_2' = d^4$ equals the volume of each grid cell. Table 3 presents results computed using (29) with $l = 12$, $d = 0.2$ (i.e., $m = 61^4$).

III. Validation of Theorem 1 through Alternative Methods

Some of these results can also be obtained through analytical integration, as detailed in Appendix A, with the relevant values listed in Table 3. It should be noted that to ensure well-defined values for D_{KL}, we select the Gaussian and Parabolic distributions, which have broader distribution ranges, as ρ_2 in the f-divergence calculation formula.

Additionally, we can arbitrarily choose a covariance matrix Σ in the (x_1, x_1', x_2, x_2') phase space and employ a grid method to compute the f-divergences. This approach discretizes the continuous phase space into regular grid cells, then constructs a numerical model based on grid nodes to approximate the difference between the two beam distributions. A schematic diagram of this method is shown in Fig. 1.

Table 3. The D_f of different types of distributions in the 4D transverse phase space with the same covariance matrix $\rho_1(x)$–$\rho_2(x)$

Methods	D_{KL}	Waterbag-Gaussian	Parabolic-Gaussian	Waterbag-Parabolic
Analytical	0.495923	0.134089	0.391326	0.407710
Numerical	0.495923	0.134089	0.391326	0.407710
Analytical	0.185837	0.054840	0.226934	0.262950
Numerical	0.185837	0.054840	0.226934	0.262950
Analytical	0.231856	0.071380	0.223872	0.309006
Numerical	0.231856	0.071380	0.223872	0.309006

As shown in Table 3, the results obtained by the three methods are essentially consistent, further validating the correctness of Theorem 1.

IV. Different f-Divergences Assign Different Weights to the Core, Tail, and Halo

When using the four f-divergences listed in Table 2 to quantify differences in beam distributions across distinct phase space regions, their distinct mathematical expressions impart unique quantification characteristics. To select the optimal f-divergence for precisely quantifying distribution differences in specific regions, we propose evaluating their proportional variations across different regions. An f-divergence exhibiting more significant proportional variation within a particular region indicates higher sensitivity to that region, thereby enabling more effective reflection of distributional differences. Our subsequent research will analyze the weight allocation of these four f-divergences across different regions for distributions with identical covariance matrices Σ but distinct types. This analysis will establish priority rankings for applying each f-divergence in specific regions. For convenience, we conduct this investigation in the $(r, \psi_1, \psi_2, \psi_3)$ phase space.

First, we establish the convention for region partitioning. Fig. 2 shows the radial projections of the density functions along coordinate r for the three beam distributions represented by Eqs. (26), (27), and (28). From Fig. 2, it can be observed that each pair of distributions shares points of equal density. The radial coordinate values of these iso-density points for each case are as follows:

- Waterbag versus Gaussian: $\sqrt{6}$
- Parabolic versus Gaussian (intersect twice): 1.68458, 2.7382
- Waterbag versus Parabolic: $\sqrt{8}$

The iso-density points and boundary points of these distributions partition the region into three parts: the beam core, tail, and halo. As shown in Fig. 3, looking outward from the coordinate origin, we define:

Core: Region within the first iso-density point between two distributions
Tail: Region between the first iso-density point and the nearest distribution boundary point
Halo: Region between the nearest and farthest distribution boundary points (i.e., the non-overlapping region between distributions)

When previously calculating the f-divergences between different distributions in the $(r, \psi_1, \psi_2, \psi_3)$ phase space using Eq. (24), we simultaneously obtained the variation curves of the four types of f-divergences (listed in Table 2) with respect to the integration radius. The integral curves of each divergence are shown in Figs. 4, 5, 6, and 7, respectively. From these figures, we can determine the weight distribution of each f-divergence in the core, tail, and halo regions—that is, the proportion of the integral value within each region to the total integral. The results are shown in Table 4. It should be noted that, due to the KL divergence curve exhibiting negative values and contributing nothing in the halo region, the KL divergence is excluded from the discussion. Its integral curve is shown in Fig. 4.

Table 4. The weights assigned to different f-divergences in different regions

Distribution Region	Waterbag-Gaussian	Parabolic-Gaussian	Waterbag-Parabolic
	JS(%)	TV(%)	Hel(%)
Core	4.6	24.5	34.0
Tail	48.5	74.5	63.3
Halo	47.0	1.0	2.7

By comparing the data in Table 4, we can determine the sensitivity of JS divergence, TV distance, and Hellinger distance to the beam core, tail, and halo regions, thereby establishing their usage priority in different regions:

- **Core:** $D_{Hel} \to D_{JS} \to D_{TV}$
- **Tail:** $D_{TV} \to D_{JS} \to D_{Hel}$
- **Halo:** $D_{JS} \to D_{TV} \to D_{Hel}$

This conclusion provides practical guidance for selecting appropriate f-divergences to quantify beam distribution differences in regions of interest.

V. The Relationship Between f-Divergences and Mismatch Factors

A. Mismatch Factor

In the 2D phase space, the three independent second-order moments in Eq. (2) are conventionally expressed using Twiss parameters: $\Sigma = \varepsilon_{rms} \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix}$ where the Twiss parameters characterize the orientation and shape of the phase space distribution, satisfying $\beta\gamma - \alpha^2 = 1$. The mismatch factor quantifies the difference between two beam distributions in 2D phase space that share the same ε_{rms} but have different Σ. Geometrically, this manifests as two RMS phase ellipses with identical centers and areas that are not fully coincident. If one ellipse can exactly enclose the other after being scaled by a certain proportion, then the mismatch factor is defined as the value of this scaling factor minus 1. It is not difficult to see from this definition that the mismatch factor measures the distribution differences between two beams in phase space through RMS boundaries. The commonly used formula is as follows [34]:

$M = \frac{\Delta + \sqrt{\Delta(\Delta+4)}}{2}$ where $\Delta = \beta_1\gamma_2 + \gamma_1\beta_2 - 2\alpha_1\alpha_2 - 2$.

B. 2D f-Divergences in Relation to Mismatch Factor

Theorem 2. In 2D phase space, the f-divergences between two beam distributions with elliptic symmetry and identical RMS emittance depends solely on their mismatch factor.

The proof of Theorem 2 can be seen in Ref. [33]. We now provide a simpler proof. In the (x_1, x_1') phase space, according to Eqs. (1) and (4), the f-divergences between two quadratic beam distributions $\rho_1(x)$ and $\rho_2(x)$ with identical RMS emittance ε_{rms} and centered at the origin is:

$$D_f[\rho_1(x)\|\rho_2(x)] = \int \rho_2(x) f\left(\frac{k_1 \cdot g_1(x^T\Sigma_1^{-1}x)}{k_2 \cdot g_2(x^T\Sigma_2^{-1}x)}\right) dx_1 dx_1'$$

From Lemma 2 in Appendix B, after applying the transformation in Eq. (B26) to the (x_1, x_1') phase space, the RMS phase ellipses of $\rho_1(x)$ and $\rho_2(x)$ can be transformed into an RMS unit phase circle and an RMS standard phase ellipse respectively in the (u_1, u_1') phase space. The relationship between phase space volume elements before and after coordinate transformation is: $dx_1 dx_1' = du_1 du_1' \cdot |J|$ where $|J| = \varepsilon_{rms}$. Thus, in the (u_1, u_1') phase space, Eq. (39) equivalently becomes:

$$D_f[\rho_1(Ru)\|\rho_2(Ru)] = \int k_2 \cdot g_2(u^T\Lambda_2^{-1}u) f\left(\frac{k_1 \cdot g_1(u^T\Lambda_1^{-1}u)}{k_2 \cdot g_2(u^T\Lambda_2^{-1}u)}\right) du_1 du_1'$$

where $\Lambda_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is the identity matrix, and $\Lambda_2 = \mathrm{diag}((M+1)^2, (M+1)^{-2})$ is a diagonal matrix defined by the mismatch factor M between $\rho_1(x)$ and $\rho_2(x)$. It is evident that once the coefficients k_1, k_2 and the functions g_1, g_2 are specified, the result of Eq. (42) depends solely on the parameter M. This indicates that for two different 2D beam distributions with elliptical symmetry, the f-divergences between them depends solely on the mismatch factor, and is independent of the orientation of the ellipsoids in phase space. Q.E.D.

C. 2n-D f-Divergences in Relation to Mismatch Factors

Mathematically, for beam distributions with elliptic symmetry that are uncoupled in the motion directions of $i = 1, 2, \ldots, n$, we can readily extend Theorem 2 to 2n-D phase space, obtaining the following theorem:

Theorem 3. In the $(x_1, x_1', x_2, x_2', \ldots, x_n, x_n')$ phase space (2n-D), given two elliptically symmetric and uncoupled beam distributions in the motion directions, if their RMS emittances in the 2D subspaces are identical, then the f-divergence between them depends solely on the mismatch factors in these 2D subspaces (x_i, x_i') for $i = 1, 2, \ldots, n$.

To further verify the correctness of this theorem, we employed the grid method mentioned in Sec. III B to simulate the relationship between the f-divergences and the two transverse mismatch factors for both same and different kinds of common ideal beam distributions (Waterbag, Parabolic, Gaussian) with elliptical symmetry and no x-y coupling in the 4D transverse phase space (x, x', y, y').

First, we selected a standard distribution and applied rotation and scaling transformations to it, ensuring that the RMS emittance of its projected distributions in the (x, x') and (y, y') phase spaces remained unchanged, as shown in Fig. 8. Subsequently, we calculated a series of mismatch factors M_x and M_y generated in these 2D subspaces by such transformations, while also computing the 4D f-divergences between the transformed distributions and the selected standard distribution.

The simulation results for the KL divergence are shown in Fig. 9. As can be seen from the figure and discussed in Sec. II B, the KL divergence is asymmetric. Additionally, for points where $\rho_1 \neq 0$ and $\rho_2 = 0$, we set ρ_2 to a very small value ($\rho_2 = 10^{-21}$) to obtain a well-defined D_{KL}, with the corresponding simulation results presented in Fig. 9. The simulation results for the JS divergence, TV distance, and Hellinger distance are shown in Figs. 10, 11, and 12, respectively. These figures demonstrate that all three divergences exhibit symmetry.

Furthermore, when $M_x = M_y = 0$, the values of the various f-divergences in these figures align with the results listed in Table 3. These figures also reveal that the curves depicting the relationship between the mismatch factors and f-divergences under both transformation methods coincide. This indicates that in the 4D transverse phase space (x, x', y, y'), for any two beam distributions with elliptical symmetry and no x-y coupling, as long as their transverse emittances are identical, the f-divergences between them depends solely on the mismatch factors M_x and M_y. This observation is consistent with the description in Theorem 3.

VI. The Relationship Between f-Divergences and Emittance Differences

A. 2D f-Divergences in Relation to Emittance Difference

Theorem 4. In the 2D phase space, the f-divergences between two matched beam distributions with elliptical symmetry depends solely on their RMS emittance scaling factor.

We now prove Theorem 4. In the (x_1, x_1') phase space, two beam distributions $\rho_1(x)$ and $\rho_2(x)$ with elliptical symmetry are said to be matched if they share identical Twiss parameters. Their covariance matrices satisfy the following relationship: $\Sigma_2 = N^2\Sigma_1$ where $N = \sqrt{\det(\Sigma_2)}/\sqrt{\det(\Sigma_1)}$ is termed the RMS emittance scaling factor between $\rho_1(x)$ and $\rho_2(x)$. The f-divergences between them can then be expressed as:

$$D_f[\rho_1(x)\|\rho_2(x)] = \int \rho_2(x) f\left(\frac{k_1 \cdot g_1(x^T\Sigma_1^{-1}x)}{k_2 \cdot g_2(x^T\Sigma_2^{-1}x)}\right) dx_1 dx_1'$$

After applying the two transformations from Eqs. (7) and (12) in Sec. III A to x, Eq. (47) equivalently becomes:

$$D_f[\rho_1(PQz)\|\rho_2(PQz)] = \int k_2 \cdot g_2(z^T(N^2I)^{-1}z) f\left(\frac{k_1 \cdot g_1(z^T I^{-1} z)}{k_2 \cdot g_2(z^T(N^2I)^{-1}z)}\right) dz_1 dz_1'$$

where I denotes the identity matrix. From Eq. (48), it is evident that if $\rho_1(x)$ and $\rho_2(x)$ are matched, their f-divergences depends solely on their RMS emittance scaling factor N. Q.E.D.

B. 2n-D f-Divergences in Relation to Emittance Differences

Mathematically, for beam distributions with elliptic symmetry that are uncoupled in the motion directions of $i = 1, 2, \ldots, n$, we can similarly generalize Theorem 4 to a 2n-D phase space, obtaining the following theorem:

Theorem 5. In the $(x_1, x_1', x_2, x_2', \ldots, x_n, x_n')$ phase space (2n-D), given two elliptically symmetric and uncoupled beam distributions in the motion directions, if the mismatch factors in these 2D subspaces are all zero, then the f-divergence between them depends solely on the RMS emittance scaling factors in the 2D subspaces represented by (x_i, x_i') for $i = 1, 2, \ldots, n$.

We similarly employed the grid method mentioned in Sec. III B to simulate the relationship between the f-divergences and the two transverse RMS emittance scaling factors for same and different kinds of common ideal beam distributions (Waterbag, Parabolic, Gaussian) with elliptical symmetry and no x-y coupling in the 4D transverse phase space (x, x', y, y'). First, we selected a standard distribution with covariance matrix equal to the identity matrix I. We then proportionally scaled this distribution in the (x, x') and (y, y') 2D subspaces according to the RMS emittance scaling factors N_x and N_y, respectively, ensuring $M_x = M_y = 0$. Simultaneously, we computed the 4D f-divergences between the scaled distribution and the selected standard distribution.

The simulation results for the KL divergence are presented in Fig. 13. For points where $\rho_1 \neq 0$ and $\rho_2 = 0$, we similarly set $\rho_2 = 10^{-21}$ during simulation to ensure D_{KL} remains well-defined. The simulation results for the JS divergence, TV distance, and Hellinger distance are shown in Figs. 14, 15, and 16, respectively. These results demonstrate that when $N_x = N_y = 1$, the values of all f-divergences in these figures align with those listed in Table 3.

Additionally, these figures reveal that apart from the KL divergence, the other three divergences exhibit a symmetry property. For the same beam distribution, if we denote the transverse RMS emittance scaling factor between the scaled distribution and the standard distribution as $N_y^{(i)}$ (where $i = 1$ indicates contraction and $i = 2$ indicates expansion), then taking $N_y = 1$ as the boundary, when the transverse RMS emittance scaling factor satisfies the following relationship:

$N_y^{(1)} \cdot N_y^{(2)} = 1$, where $N_y^{(1)}, N_y^{(2)} \in (0, \infty)$

the values of D_{JS}, D_{TV}, and D_{Hel} are equal. For different beam distributions, if $N_y^{(i)}$ and $\hat{N}_y^{(i)}$ represent the RMS transverse emittance scaling factors corresponding to $D_f(\rho_1 \| \rho_2)$ and $D_f(\rho_2 \| \rho_1)$ respectively, then taking $N_y = 1$ as the boundary, when the RMS transverse emittance scaling factors satisfy the following relationship:

$$N_y^{(1)} \cdot \hat{N}_y^{(2)} = \hat{N}_y^{(1)} \cdot N_y^{(2)} = 1$$

the values of D_{JS}, D_{TV}, and D_{Hel} are equal.

VII. Assessment Standards for f-Divergences

In the (x, x', y, y') phase space, for ideal beam distributions exhibiting elliptical symmetry without x-y coupling, when the transverse RMS emittance scaling factors satisfy $N_x = N_y = 1$, we can establish the first assessment standard for 4D f-divergences based on the relationship between f-divergences and mismatch factors described in Theorem 3. Figs. 17, 18, and 19 present assessment heatmaps for the same distributions, while Figs. 20, 21, and 22 provide assessment heatmaps for different distributions. In all figures, the ranges of M_x and M_y are $[0, 1]$.

When the transverse mismatch factors satisfy $M_x = M_y = 0$, we can establish the second assessment standard for 4D f-divergences based on the relationship between f-divergences and RMS emittance scaling factors described in Theorem 5. Figs. 23, 24, 25, and 26 present assessment heatmaps for N_x and N_y in the range $[0.1, 1]$. Additionally, for f-divergences excluding KL divergence, using the symmetries described in Eqs. (49) and (50) for same and different distribution types respectively, we can directly map computational results from the $[0.1, 1]$ interval to the $[1, 10]$ interval.

The value ranges of f-divergences are shown in Table 2.

When quantifying differences between ideal beam distributions using f-divergences, we previously could only infer that values closer to 0 indicated smaller differences while those approaching the upper limit suggested larger differences. However, we could not interpret their concrete physical meaning or precisely gauge the degree of difference. With these assessment heatmaps, we can now analyze how much transverse mismatch factors or RMS emittance scaling factors contribute to observed differences under two specific conditions ($N_x = N_y = 1$ or $M_x = M_y = 0$), thereby enabling difference traceability.

[Figure 23: see original paper]. The Second Assessment Heatmap for 4D Kullback-Leibler Divergence (a) Waterbag versus Waterbag (b) Waterbag versus Parabolic (c) Waterbag versus Gaussian (d) Parabolic versus Waterbag (e) Parabolic versus Parabolic (f) Parabolic versus Gaussian (g) Gaussian versus Waterbag (h) Gaussian versus Parabolic (i) Gaussian versus Gaussian

[Figure 24: see original paper]. The Second Assessment Heatmap for 4D Jensen-Shannon Divergence (a) Waterbag versus Waterbag (b) Waterbag versus Parabolic (c) Waterbag versus Gaussian (d) Parabolic versus Waterbag (e) Parabolic versus Parabolic (f) Parabolic versus Gaussian (g) Gaussian versus Waterbag (h) Gaussian versus Parabolic (i) Gaussian versus Gaussian

[Figure 25: see original paper]. The Second Assessment Heatmap for 4D Total Variation Distance (a) Waterbag versus Waterbag (b) Waterbag versus Parabolic (c) Waterbag versus Gaussian (d) Parabolic versus Waterbag (e) Parabolic versus Parabolic (f) Parabolic versus Gaussian (g) Gaussian versus Waterbag (h) Gaussian versus Parabolic (i) Gaussian versus Gaussian

[Figure 26: see original paper]. The Second Assessment Heatmap for 4D Hellinger Distance (a) Waterbag versus Waterbag (b) Waterbag versus Parabolic (c) Waterbag versus Gaussian (d) Parabolic versus Waterbag (e) Parabolic versus Parabolic (f) Parabolic versus Gaussian (g) Gaussian versus Waterbag (h) Gaussian versus Parabolic (i) Gaussian versus Gaussian

3. Some of these results can also be obtained through analytical integration (continued)

In other words, these two assessment standards provide calibration scales for f-divergences that previously only had measurement ranges, helping clarify the physical significance of f-divergence values.

These standards can also provide initial guidance for assessing differences between non-ideal beam distributions.

Specifically, we first use f-divergences to determine which ideal distribution type (Waterbag, Parabolic, or Gaussian) each beam approximates. Then we refer to the corresponding assessment heatmaps for analysis.

It should be noted that assessment standards for KL divergence are not uniquely defined. This non-uniqueness primarily stems from two aspects: First, when beam distributions share identical RMS emittances, D_{KL} exhibits the asymmetry described by Eq. (5). Conversely, when beams are matched, D_{KL} lacks the symmetries defined in Eqs. (49) and (50). Second, when processing points where $\rho_1 \neq 0$ and $\rho_2 = 0$, we set ρ_2 to a small constant c (typically $c = 10^{-6}$ in this work) to ensure D_{KL} yields a defined value (see Sec. IIB, Case 3 on D_{KL}'s domain). However, D_{KL} values depend critically on the choice of c. Evidently, for practical use and intuitive interpretation in beam distribution assessment, KL divergence proves notably less convenient than the other three divergences listed in Table 2.

Future work will focus on beam distributions with coupling in the direction of motion.

Appendix A: Analytical Integral Expressions of Some f-Divergences

In this section, to simplify the presentation, we adopt the following equivalent notation conventions:

$$\int_R \int_0^{\pi} \int_0^{\pi} \int_0^{2\pi} dV_4 \equiv \int r^3 \sin^2 \psi_1 \sin \psi_2 \, dr \, d\psi_1 \, d\psi_2 \, d\psi_3$$

VIII. CONCLUSION

This work presents the first systematic study of quantification characteristics for four representative f-divergences (D_{KL}, D_{JS}, D_{TV}, and D_{Hel}) in common ideal high-dimensional beam distributions. We first prove that f-divergences between different beam distribution types sharing the same covariance matrix Σ are independent of second-order moment values within Σ. Using three distinct computational methods, we calculate 4D f-divergence values under this condition. Based on these findings, we further establish the regional sensitivities of D_{JS}, D_{TV}, and D_{Hel} toward beam core, tail, and halo regions. Subsequently, we demonstrate the correspondence between f-divergences and both mismatch factors and RMS emittance scaling factors for uncoupled beam distributions in all motion directions. We validate this correspondence in (x, x', y, y') phase space by computing f-divergences between identical and different beam distributions across varying mismatch factors and RMS emittance scaling factors, while simultaneously testing divergence applicability to common beam distributions. Finally, based on this correspondence, we establish two assessment standards for f-divergences to facilitate physical interpretation within beam phase space distributions. These findings not only provide guidance for selecting f-divergences to quantify differences in high-dimensional beam distributions, but also help effectively trace the sources of differences after quantification.

Appendix B: Several Lemmas

Lemma 1

For any 2n-D quadratic form distribution $\rho(x) = kg(x^T \Sigma^{-1} x)$, there always exists an orthogonal transformation $x = Py$ that reduces $\rho(x)$ to the standard form distribution $\rho(Py) = kg(y^T \Lambda^{-1} y)$, where Λ is a diagonal matrix whose diagonal elements are the 2n positive eigenvalues of Σ.

Proof: Since Σ is a symmetric and positive definite matrix, there exists an orthogonal transformation matrix P such that $P^T \Sigma P = \Lambda = \mathrm{diag}(\lambda_1, \lambda_1', \ldots, \lambda_n, \lambda_n')$, where $\lambda_1, \lambda_1', \ldots, \lambda_n, \lambda_n'$ are the positive eigenvalues of Σ. The transformation converts $d^2 = x^T \Sigma^{-1} x$ into $d^2 = (Py)^T \Sigma^{-1} (Py) = y^T (P^T \Sigma P)^{-1} y = y^T \Lambda^{-1} y$. Thus, after the transformation, the distribution becomes $\rho(Py) = kg(y^T \Lambda^{-1} y)$. Geometrically, this represents a 2n-D standard ellipsoidal distribution centered at the origin with semi-axes parallel to the coordinate axes. Q.E.D.

Lemma 2

Given any two phase ellipses with identical emittance but different Twiss parameters, there always exists a transformation matrix R that reduces them to a unit phase circle and a standard phase ellipse, respectively.

Proof: As shown in Fig. 28(a), consider two phase ellipses in the (x_1, x_1') phase space given by $x^T \Sigma_1^{-1} x = 1$ and $x^T \Sigma_2^{-1} x = 1$, with covariance matrices: $\Sigma_1 = \varepsilon \begin{pmatrix} \beta_1 & -\alpha_1 \\ -\alpha_1 & \gamma_1 \end{pmatrix}$ and $\Sigma_2 = \varepsilon \begin{pmatrix} \beta_2 & -\alpha_2 \\ -\alpha_2 & \gamma_2 \end{pmatrix}$.

First, apply the transformation $x = R_1 m$, where $R_1 = \sqrt{\varepsilon} \begin{pmatrix} \sqrt{\beta_1} & 0 \\ -\alpha_1/\sqrt{\beta_1} & 1/\sqrt{\beta_1} \end{pmatrix}$. After this transformation, Eq. (B8) becomes $m^T \sigma_1^{-1} m = 1$, where $\sigma_1 = R_1^{-1} \Sigma_1 (R_1^{-1})^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. In the (m_1, m_1') phase space, this represents a unit phase circle.

Similarly, Eq. (B9) becomes $m^T \sigma_2^{-1} m = 1$, where $\sigma_2 = R_1^{-1} \Sigma_2 (R_1^{-1})^T$. This represents a phase ellipse with a tilt angle θ.

Next, apply the rotation transformation $m = R_2 u$, where $R_2 = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. After this transformation, Eq. (B14) becomes $u^T \Lambda_1^{-1} u = 1$, where $\Lambda_1 = R_2^{-1} \sigma_1 (R_2^{-1})^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Similarly, Eq. (B16) becomes $u^T \Lambda_2^{-1} u = 1$, where $\Lambda_2 = R_2^{-1} \sigma_2 (R_2^{-1})^T = \mathrm{diag}((M+1)^2, (M+1)^{-2})$. Here M is the mismatch factor between the two phase ellipses.

In the (u_1, u_1') phase space, the first equation still represents a unit phase circle, while the second represents a standard phase ellipse, as shown in Fig. 28(c). Therefore, for any two phase ellipses with identical emittance but different Twiss parameters, there always exists a transformation matrix $R = R_1 R_2$ that reduces them to a unit phase circle and a standard phase ellipse, respectively. Q.E.D.

References

[1] Z.J. Wang, S.H. Liu, W.L. Chen et al., Beam physics design of a superconducting linac. Phys. Rev. Accel. Beams 27, 010101 (2024). doi: 10.1103/PhysRevAccelBeams.27.010101

[2] S.H. Liu, Z.J. Wang, H. Jia et al., Physics design of the CIADS 25MeV demo facility. Nucl. Instrum. Methods Phys. Res. A 843, 11-17 (2017). doi: 10.1016/j.nima.2016.10.055

[3] W.P. Dou, Z.J. Wang, F.J. Jia et al., Beam dynamic design of a high intensity injector for proton linac. Nucl. Instrum. Methods Phys. Res. A 827, 145-151 (2016). doi: 10.1016/j.nima.2016.04.108

[4] K.R. Crandall and D.P. Rusthoi, Trace 3-d documentation. (1987). doi: 10.2172/6290515

[5] P. Forck, Lecture notes on beam instrumentation and diagnostics. https://indico.cern.ch/event/1470062/contributions/6188910/attachments/3024028/5336512/juas_script. (2025).

[6] J.C. Wong, S.M. Lund, and T. Maruta, High resolution phase space measurements with Allison-type emittance scanners. Phys. Rev. Accel. Beams 25, 064201 (2019). doi: 10.1103/PhysRevAccelBeams.22.072801

[7] E. Prat and M. Aiba, Four-dimensional transverse beam matrix measurement using the multiple-quadrupole scan technique. Phys. Rev. Accel. Beams 17, 052801 (2014). doi: 10.1103/PhysRevSTAB.17.052801

[8] C. Xiao, L. Groening, P. Gerhard et al., Measurement of the transverse four-dimensional beam rms-emittance of an intense uranium beam at 11.4 MeV/u. Nucl. Instrum. Methods Phys. Res. A 820, (2016). doi: 10.1016/j.nima.2016.02.090

[9] C. Xiao, M. Maier, X. N. Du et al., Rotating system for four-dimensional transverse rms-emittance measurements. Phys. Rev. Accel. Beams 19, 072802 (2016). doi: 10.1103/physrevaccelbeams.19.072802

[10] A. Wolski, D. C. Christie, B. L. Militsyn et al., Transverse phase space characterization in an accelerator test facility. Phys. Rev. Accel. Beams 23, 032804 (2020). doi: 10.1103/physrevaccelbeams.23.032804

[11] H. Kremers, J. Beijers, S. Brandenburg, A pepper-pot emittance meter for low-energy heavy-ion beams. Rev. Sci. Instrum. 84, (2013). 10.1063/1.4793375

[12] C.E. Mitchell, R.D. Ryne, K. Hwang, Using kernel-based statistical distance to study the dynamics of charged particle beams in particle-based simulation codes. Phys. Rev. E 106, 065302 (2022). doi: 10.1103/PhysRevE.106.065302

[13] S. Kullback and R. Leibler, On Information and Sufficiency. Annals of Mathematical Statistics 22, 79-86 (1951). doi: 10.1214/AOMS/1177729694

[14] D.I. Belov, R.D. Armstrong, Distributions of the Kullback-Leibler divergence with applications. British Journal of Mathematical & Statistical Psychology 64, 291-309 (2011). doi: 10.1348/000711010X522227

[15] J. Lin, Divergence measures based on the Shannon entropy. IEEE Trans. Inf. Theory 37, 145-151 (1991). doi: 10.1109/18.61115

[16] F. Nielsen, On a Variational Definition for the Jensen-Shannon Symmetrization of Distances based on the Information Radius. Entropy 23, (2021). doi:10.3390/e23040464

[17] X. Li, L.H. Xu, H. Yang, Stable Central Limit Theorem in Total Variation Distance. J Theor Probab 38, (2025). doi: 10.1007/s10959-024-01385-7

[18] G.L. Gilardoni, On the Minimum f-Divergence for Given Total Variation. C. R. Math. Acad. Sci. Paris 343, 763-766 (2006). doi: 10.1016/j.crma.2006.10.027

[19] E. Hellinger, Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. J. Reine Angew. Math. 136, 210-271 (1909). doi: 10.1515/crll.1909.136.210

[20] D.A. Cieslak T.R. Hoens, N.V. Chawla et al., Hellinger distance decision trees are robust and skew-insensitive. Data Mining Knowledge Discovery 24, 136-158 (2012). doi: 10.1007/s10618-011-0222-1

[21] A. Hoover and J.C. Wong, High-dimensional maximum-entropy phase space tomography using normalizing flows. Phys. Rev. Research 6, 033163 (2024). doi: 10.1103/PhysRevResearch.6.033163

[22] J.C. Wong, A. Shishlo, A. Aleksandrov et al., 4D transverse phase space tomography of an operational hydrogen ion beam via noninvasive 2D measurements using laser wires. Phys. Rev. Accel. Beams 25, 042801 (2022). doi: 10.1103/PhysRevAccelBeams.25.042801

[23] G.N. Minerbo, O.R. Sander, and R.A. Jameson, Four-dimensional beam tomography. IEEE Trans. Nucl. Sci. 28, 2231 (1981). doi: 10.1109/TNS.1981.4331646

[24] C.T. Mottershead, Maximum entropy beam diagnostic tomography. IEEE Trans. Nucl. Sci. 32, 1970 (1985). doi: 10.1109/TNS.1985.4333784

[25] K. Hock and A. Wolski, Tomographic reconstruction of the full 4D transverse phase space. Nucl. Instrum. Methods Phys. Res. A 726, 8 (2013). doi: 10.1016/j.nima.2013.05.004

[26] M.W. Wang, Z.M. Wang, D. Wang et al., Four-dimensional phase space measurement using multiple two-dimensional profiles. Nucl. Instrum. Methods Phys. Res. A 943, 162438 (2019). doi: doi.org/10.1016/j.nima.2019.162438

[27] L.W. Liu, Z.J. Wang, J.C. Wong et al., Maximum entropy tomography of 4D transverse phase space distribution using 2D measurement results. in Proc. IBIC2024, Beijing, 166-169 (2024). doi: 10.18429/JACoW-IBIC2024-TUP51

[28] L.W. Liu, Z.J. Wang, J.C. Wong et al., Maximum entropy phase space tomography under nonlinear beam transport. in Proc. LINAC2024, Chicago, IL, USA, 438-441 (2024). doi: 10.18429/JACoW-LINAC2024-TUPB057

[29] M. Reiser, Theory and Design of charged Particle Beams. (2008). doi: 10.1002/9783527622047

[30] V.V. Prelov, The f-Divergence and Coupling of Probability Distributions. Probl Inf Transm 57, 54-69 (2021). doi: 10.1134/S0032946021010038

[31] Y. Cai and L.H. Lim, Distances Between Probability Distributions of Different Dimensions. IEEE Trans. Inf. Theory 68, 4020-4031 (2022). doi: 10.1109/TIT.2022.3148923

[32] R. Xie and W. Wei, f-Divergence Distributionally Robust Optimization. In: Distributionally Robust Optimization and its Applications in Power System Energy Storage Sizing. Springer Nature Singapore, (2024), pp. 171-211. doi: 10.1007/978-981-97-2566-3_5

[33] Y. Du, Z.J. Wang, J.C. Wong et al., A comparison of RMS moments and statistical divergences as ways to quantify the difference between beam phase space distributions. (2024).

Note: Figure translations are in progress. See original paper for figures.