Yu Du
duyu@impcas.ac.cn
Institute of Modern Physics,
Chinese Academy of Sciences

September 11, 2024

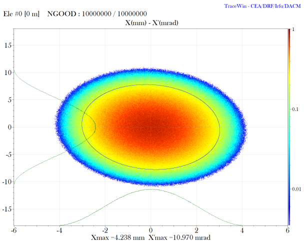
Evaluating the Use of Common Statistical Divergences to Quantify the Differences Between Beam Distributions in High-Dimensional Phase Space

The 13th International Beam Instrumentation Conference (IBIC2024)

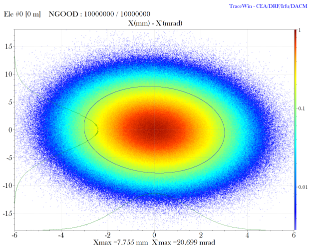
Contents

Background

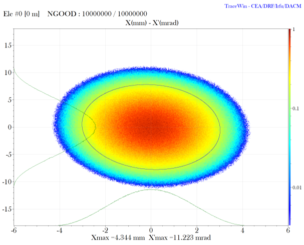
Different types of distributions with the same RMS moments



Parabolic



6 standard deviations Gaussian



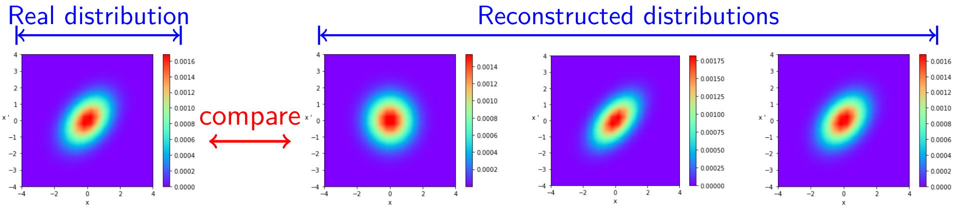
Water Bag

Clear differences exist between distributions, how to quantify?

- Method 1: Using RMS moments → No difference? ✗
- Method 2: Comparing distribution differences point by point. ✓

Background

Beam 4D transverse phase space tomography



(x, x') *projection* — *2 profiles* — *4 profiles* — *6 profiles*

- As more profiles are selected, the reconstructed distribution increasingly approaches the real distribution.
- How to measure the degree of difference between them?

Demands: Accurately quantifying the difference between two beam distributions in high-dimensional phase space.

Significance: It's crucial for the interpretation of experimental and simulation results.

Methods: Using statistical divergences.

f-divergences

- The f-divergences are a common class of methods used to measure the difference between two probability distributions, defined as follows:

$$D_f[p(\mathbf{x})||q(\mathbf{x})] := \int q(\mathbf{x}) f\left[\frac{p(\mathbf{x})}{q(\mathbf{x})}\right] \mathrm{d}^n\mathbf{x}$$

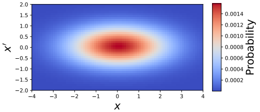
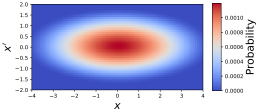
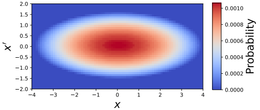
- $f(\cdot)$ is a convex function and satisfies $f(1) = 0$;
- Different $f(\cdot)$ correspond to different statistical divergences:

Name	$f(t)$	$D_f[p(\mathbf{x})\|\|q(\mathbf{x})]$
Kullback-Leibler	$t \ln t$	$\int p(\mathbf{x}) \ln\left[\frac{p(\mathbf{x})}{q(\mathbf{x})}\right] \mathrm{d}^n\mathbf{x}$
Jensen-Shannon	$\frac{1}{2}\left[(t+1)\ln\left(\frac{2}{t+1}\right) + t\ln t\right]$	$\frac{1}{2}\int \left\{q(\mathbf{x}) \ln\left[\frac{2q(\mathbf{x})}{p(\mathbf{x})+q(\mathbf{x})}\right] + p(\mathbf{x}) \ln\left[\frac{2p(\mathbf{x})}{p(\mathbf{x})+q(\mathbf{x})}\right]\right\} \mathrm{d}^n\mathbf{x}$
Total Variation	$\frac{1}{2}\|t-1\|$	$\frac{1}{2}\int \|p(\mathbf{x}) - q(\mathbf{x})\| \mathrm{d}^n\mathbf{x}$
Squared Hellinger	$(\sqrt{t}-1)^2$	$\int \left[\sqrt{p(\mathbf{x})} - \sqrt{q(\mathbf{x})}\right]^2 \mathrm{d}^n\mathbf{x}$

Four forms of f-divergences; $x \in \mathbb{R}^n$, $t = p(\mathbf{x})/q(\mathbf{x})$

f-divergences give the total contribution of differences at all points.

Several Common 4D Beam Distributions

Distribution with Elliptical Symmetry	Definition $\rho(x,x',y,y') = \rho(I)$ $I = \mathbf{x}^T\Sigma^{-1}\mathbf{x}$	Schematic Diagram of 2D Projection (x,x')
Gaussian	$\frac{1}{(\sqrt{2\pi})^4\lvert\Sigma\rvert^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}I}$	
Parabolic	$\frac{6}{(\sqrt{8\pi})^4\lvert\Sigma\rvert^{\frac{1}{2}}} \cdot \left(1-\frac{I}{8}\right), \quad I < 8$	
Water Bag	$\frac{2}{(\sqrt{6\pi})^4\lvert\Sigma\rvert^{\frac{1}{2}}}, \quad I < 6$	

Quadratic form expression of beam distributions with elliptical symmetry; $\mathbf{x} = (x,x',y,y')^T$;
Σ *is a covariance matrix composed of 10 independent second-order moments.*

The D_f Between Distributions of Different Types

- ρ_1 and ρ_2 have the same Σ, but different distribution types
  - The beam distributions are matched: $M_x = M_y = 0$
  - The RMS emittance is the same: $\frac{\varepsilon_x}{\varepsilon_{x0}} = \frac{\varepsilon_y}{\varepsilon_{y0}} = 1$
- We have directly calculated the divergence values between the following distributions using mathematical integration:

$\rho_1(\mathbf{x}),\ \rho_2(\mathbf{x})$	D_{KL}	D_{JS}	D_{TV}	D_{Hel}
Parabolic, Gaussian	0.185837	0.054823	0.226909	0.262902
Water Bag, Gaussian	0.495922	0.134071	0.391299	0.407679
Water Bag, Parabolic	0.231856	0.071380	0.223872	0.309006

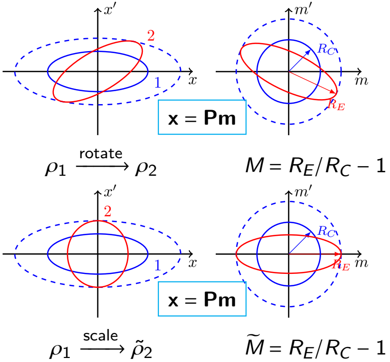
The f-divergences between different distributions with the same RMS emittance.

- For two different types of beam distributions, as long as their Σ are the same (regardless of the values), the D_f between them is fixed.

f-divergences in Relation to Mismatch Factors

Theorem

In the $(x_1, x'_1, x_2, x'_2, \cdots, x_i, x'_i, \cdots, x_n, x'_n)$ *phase space (2n-D), given two uncoupled beam distributions with elliptical symmetry, the f-divergences between them are uniquely determined by the mismatch factors in the 2D subspaces represented by the elements of the set* $\{(x_i, x'_i) \mid i = 1, 2, \cdots, n\}$.

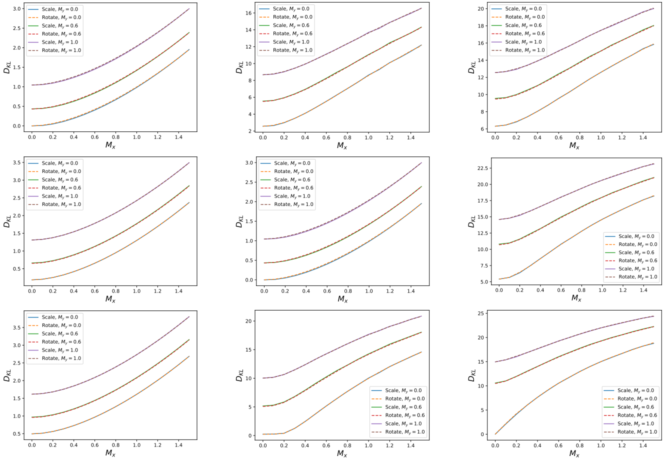


Verification Method:

- Generate the same mismatch factor in two different ways: Rotation, Scaling.
- Ensure the same ε_{rms} in the (x, x') and (y, y') phase spaces, but with second-order moments that are not completely identical.
- Check whether f-divergences obtained from these two methods is identical.

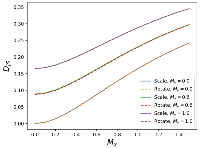
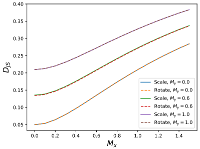
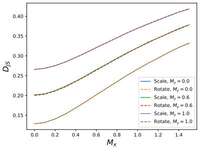
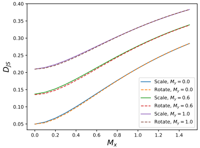
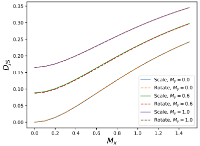
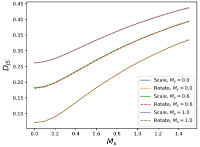
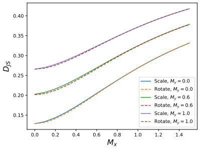
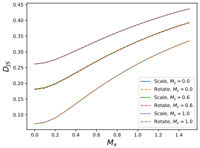
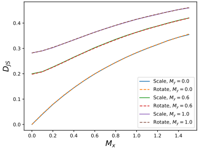
Simulation Verification | $(M_x, M_y) \mapsto D_{KL}$

KL: $\rho_1(\mathbf{x})$ \ $\rho_2(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			



The relationship between 4D KL-divergence and transverse mismatch factor(Asymmetric)

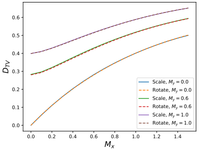
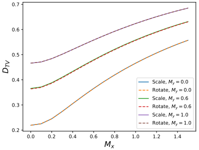
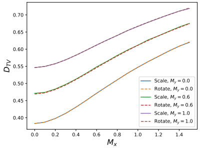
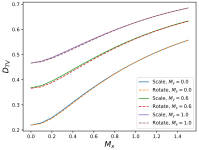
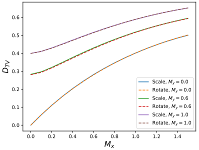
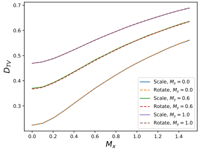
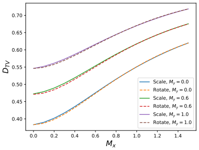
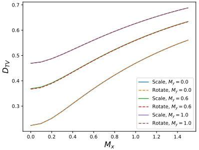
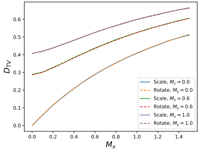
Simulation Verification | $(M_x, M_y) \mapsto D_{JS}$

JS: $\rho_1(\mathbf{x})$ \ $\rho_2(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D JS-divergence and transverse mismatch factor(Symmetry)

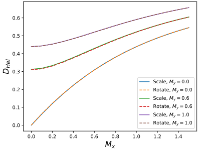
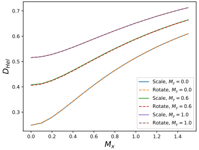
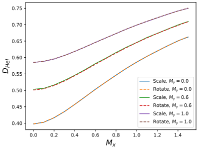
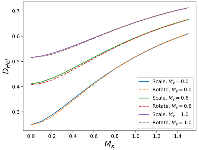
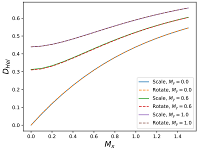
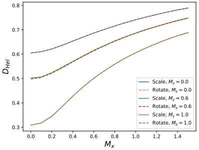
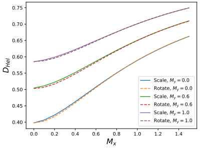
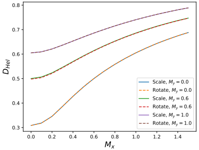
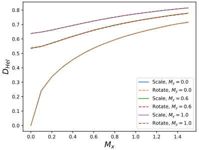
China Initiative Accelerator Driven System

Simulation Verification | $(M_x, M_y) \mapsto D_{TV}$

TV: $\rho_1(\mathbf{x})$ \\ $\rho_2(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D TV-distance and transverse mismatch factor(Symmetry)

Simulation Verification | $(M_x, M_y) \mapsto D_{Hel}$

Hel ($\rho_1(\mathbf{x})$ \ $\rho_2(\mathbf{x})$)	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D Hellinger-distance and transverse mismatch factor(Symmetry)

Discussion of Simulation Results | $(M_x, M_y) \mapsto D_f$

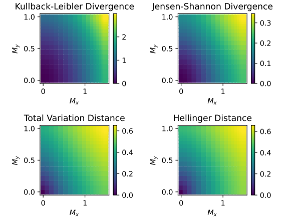
- The above figures illustrate how f-divergences vary with mismatch factors.
- The curves obtained from the two methods coincide (corresponding to the solid and dashed lines in the figure), indicating:

> **The relationship between 4D f-divergence and transverse mismatch factors**
>
> In the (x, x', y, y') phase space, for two beam distributions with elliptical symmetry and no x-y coupling, the f-divergences between them are uniquely determined by the two transverse mismatch factors.

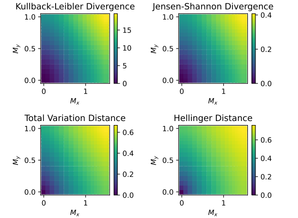
This relationship can be utilized to provide an assessment standard for these popular divergences.

Assessment Heatmap for 4D f-divergences $|(M_x, M_y) \mapsto D_f$

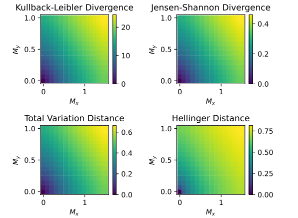
Same Distributions



Gaussian versus Gaussian

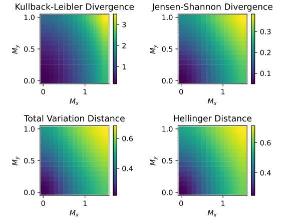


Parabolic versus Parabolic

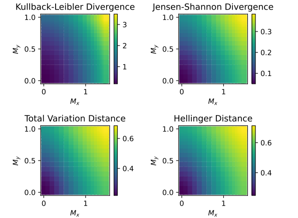


Water Bag versus Water Bag

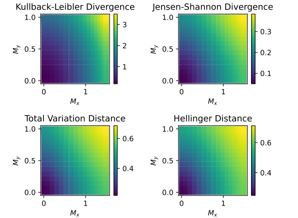
Different Distributions



Gaussian versus Parabolic



Gaussian versus Water Bag



Water Bag versus Parabolic

f-divergences in Relation to RMS Emittance

Theorem

In the $(x_1, x_1', x_2, x_2', \cdots, x_i, x_i', \cdots, x_n, x_n')$ phase space (2n-D), given two beam distributions with elliptical symmetry and no coupling, if the mismatch factors $\{M_i \mid i = 1, 2, \cdots, n\}$ in the 2D subspaces represented by $\{(x_i, x_i') \mid i = 1, 2, \cdots, n\}$ are all zero, then the f-divergences between them depend only on the scaling ratios of the RMS emittances $\{\varepsilon_i/\varepsilon_{i0} \mid i = 1, 2, \cdots, n\}$ in these 2D subspaces.

Simulation Content

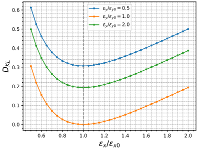
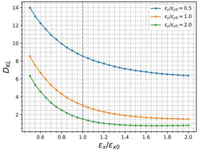
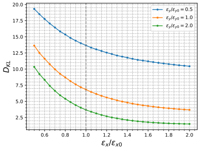
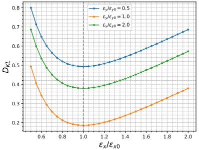
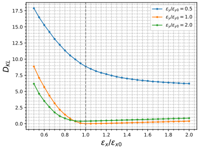
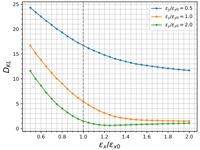
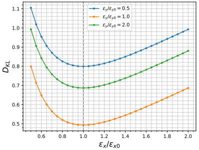
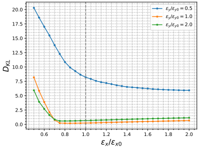
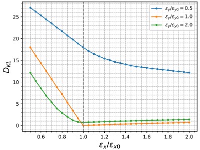
Premise: The beam distributions are matched: $M_x = M_y = 0$

Operation: Vary the RMS emittance in the (x, x') and (y, y') sub-phase spaces:

$$\varepsilon_x = k_1 \varepsilon_{x0}, \quad \varepsilon_y = k_2 \varepsilon_{y0}, \quad \varepsilon_{4D} = k_1 k_2 \cdot \varepsilon_{x0} \varepsilon_{y0}$$

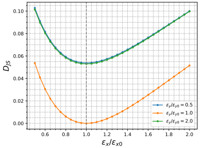
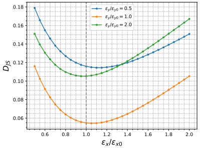
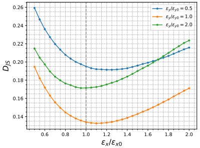
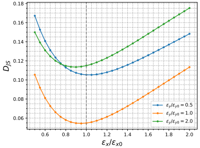
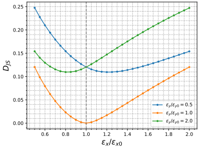
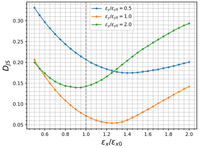
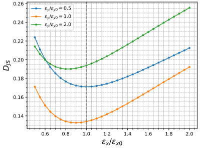
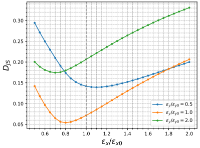
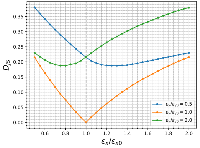
Goal: Investigate how the f-divergences vary with the $\frac{\varepsilon_x}{\varepsilon_{x0}}$ and $\frac{\varepsilon_y}{\varepsilon_{y0}}$.

Simulation Results | $\left(\frac{\varepsilon_x}{\varepsilon_{x0}}, \frac{\varepsilon_y}{\varepsilon_{y0}}\right) \mapsto D_{KL}$

KL: $\rho_1(\mathbf{x})$ \ $\rho_2(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D KL-divergence and RMS emittance

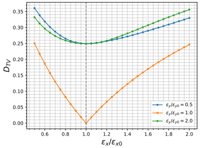
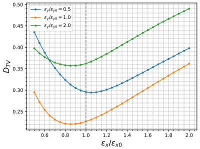
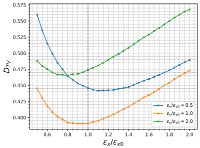
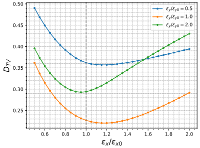
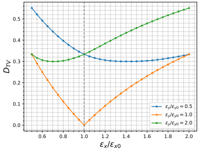
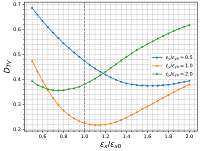
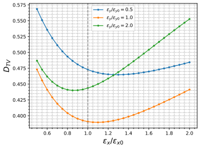
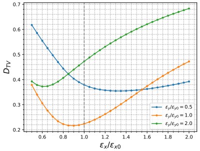
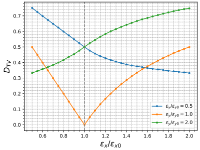
Simulation Results | $\left(\frac{\varepsilon_x}{\varepsilon_{x0}}, \frac{\varepsilon_y}{\varepsilon_{y0}}\right) \mapsto D_{JS}$

JS — $\rho_2(\mathbf{x})$ / $\rho_1(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D JS-divergence and RMS emittance

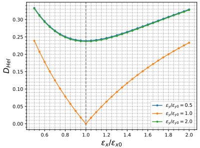
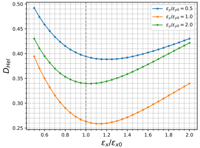
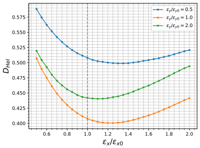
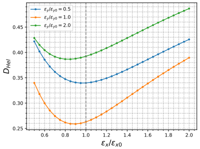
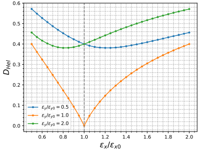
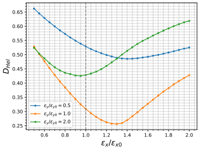
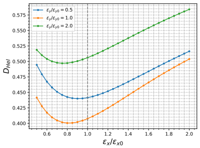
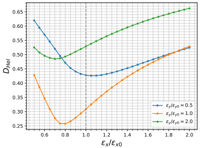
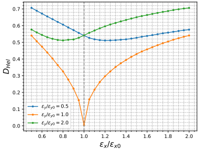
China Initiative Accelerator Driven System

Simulation Results | $(\frac{\varepsilon_x}{\varepsilon_{x0}}, \frac{\varepsilon_y}{\varepsilon_{y0}}) \mapsto D_{TV}$

TV: $\rho_1(\mathbf{x})$ \ $\rho_2(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D TV-distance and RMS emittance

Simulation Results | $\left(\frac{\varepsilon_x}{\varepsilon_{x0}}, \frac{\varepsilon_y}{\varepsilon_{y0}}\right) \mapsto D_{Hel}$

Hel $\rho_2(\mathbf{x})$ / $\rho_1(\mathbf{x})$	Gaussian	Parabolic	Water Bag
Gaussian			
Parabolic			
Water Bag			

The relationship between 4D Hellinger-divergence and RMS emittance

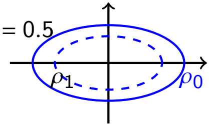
China Initiative Accelerator Driven System

Discussion of Simulation Results | $(\frac{\varepsilon_x}{\varepsilon_{x0}}, \frac{\varepsilon_y}{\varepsilon_{y0}}) \mapsto D_f$

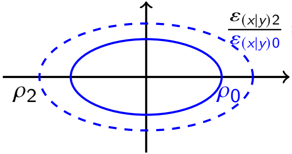
IMP

- The above figures illustrate how f-divergences vary with ${}^{\varepsilon_x}/_{\varepsilon_{x0}}$ and ${}^{\varepsilon_y}/_{\varepsilon_{y0}}$.
- Using ${}^{\varepsilon_{(x|y)}}/_{\varepsilon_{(x|y)0}} = 1$ as the boundary, D_{JS}, D_{TV} and D_{Hel} satisfy the following symmetry(values are the same), but D_{KL} does not:
  - Distributions of the same type:

$\frac{\varepsilon_{(x|y)1}}{\varepsilon_{(x|y)0}} = 0.5$



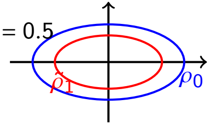
$$\begin{cases} \frac{\varepsilon_{x0}}{\varepsilon_{x1}} = \frac{\varepsilon_{x2}}{\varepsilon_{x0}} = k_x, & k_x \geq 1 \\ \frac{\varepsilon_{y0}}{\varepsilon_{y1}} = \frac{\varepsilon_{y2}}{\varepsilon_{y0}} = k_y, & k_y \geq 1 \end{cases}$$



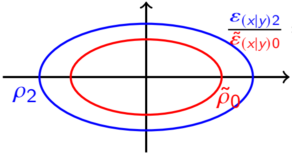
$\frac{\varepsilon_{(x|y)2}}{\varepsilon_{(x|y)0}} = 2$

  - Distributions of different type($\varepsilon_{(x|y)0} = \tilde{\varepsilon}_{(x|y)0}$):

$\frac{\tilde{\varepsilon}_{(x|y)1}}{\varepsilon_{(x|y)0}} = 0.5$



$$\begin{cases} \frac{\varepsilon_{x0}}{\tilde{\varepsilon}_{x1}} = \frac{\varepsilon_{x2}}{\tilde{\varepsilon}_{x0}} = \tilde{k}_x, & \tilde{k}_x \geq 1 \\ \frac{\varepsilon_{y0}}{\tilde{\varepsilon}_{y1}} = \frac{\varepsilon_{y2}}{\tilde{\varepsilon}_{y0}} = \tilde{k}_y, & \tilde{k}_y \geq 1 \end{cases}$$

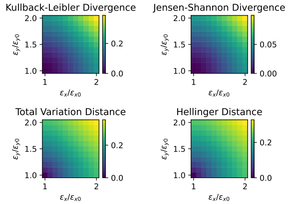


$\frac{\varepsilon_{(x|y)2}}{\tilde{\varepsilon}_{(x|y)0}} = 2$

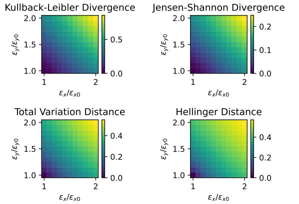
- When $M_x = M_y = 0$, the f-divergences depend solely on $\frac{\varepsilon_x}{\varepsilon_{x0}}$ and $\frac{\varepsilon_y}{\varepsilon_{y0}}$. This relationship can be utilized to provide a second assessment standard for the f-divergence.

Assessment Heatmap for 4D f-divergences $|(\frac{\varepsilon_x}{\varepsilon_{x0}}, \frac{\varepsilon_y}{\varepsilon_{y0}}) \mapsto D_f$

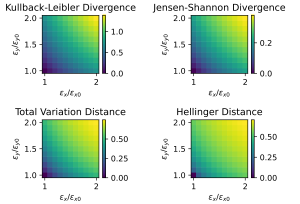
Same Distributions



Gaussian versus Gaussian

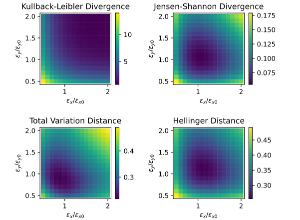


Parabolic versus Parabolic

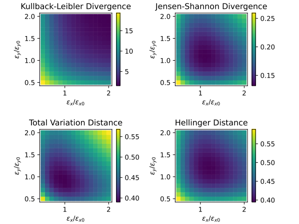


Water Bag versus Water Bag

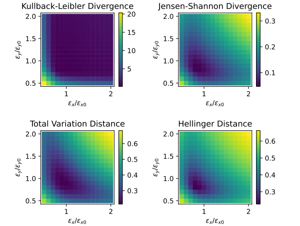
Different Distributions



Gaussian versus Parabolic



Gaussian versus Water Bag



Water Bag versus Parabolic

China Initiative Accelerator Driven System

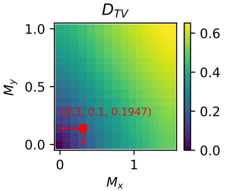
Standard Values

- We have set assessment standards using ideal distributions.
- This can offer a rough D_f assessment reference for non-ideal distributions.
  - $\boxed{\rho_{non-ideal}} \xleftrightarrow[\text{closest to}]{D_{f,min}} \boxed{\rho_{ideal}}$ (Gaussian? Parabolic? Water Bag?)
  - Corresponding to the existing evaluation heatmaps.

For example: Beam 4D transverse phase space tomography result analysis

ρ_1 closest to ↓ Gaussian — **Real distribution** ⟷ $D_{TV} = 0.1947$ ⟷ **Reconstructed distribution** — ρ_2 ↓ closest to Gaussian

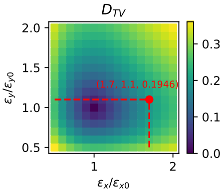
Case 1 (Gaussian-Gaussian):



$$\boxed{\frac{\varepsilon_x}{\varepsilon_{x0}} = \frac{\varepsilon_y}{\varepsilon_{y0}} = 1}$$

$$M_x = 0.3(or\ 0.1)$$
$$M_y = 0.1(or\ 0.3)$$

Case 2 (Gaussian-Gaussian):



$$\boxed{M_x = M_y = 0}$$

$$\frac{\varepsilon_x}{\varepsilon_{x0}} = 1.7(or\ \frac{1}{1.7})$$
$$\frac{\varepsilon_y}{\varepsilon_{y0}} = 1.1(or\ \frac{1}{1.1})$$

China Initiative Accelerator Driven System

Discussion on Kullback-Leibler Divergence

$$\boxed{D_{KL}\left[\rho_1||\rho_2\right] := \int \rho_1(\mathbf{x}) \ln\left[\frac{\rho_1(\mathbf{x})}{\rho_2(\mathbf{x})}\right] \mathrm{d}^n\mathbf{x}}$$

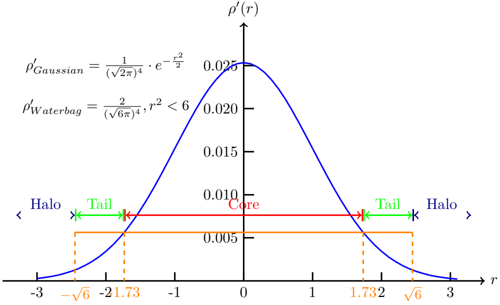
- Asymmetric
  - $\frac{\varepsilon_x}{\varepsilon_{x0}} = \frac{\varepsilon_y}{\varepsilon_{y0}} = 1$: $D_{KL}\left[\rho_1||\rho_2\right] \neq D_{KL}\left[\rho_2||\rho_1\right]$, violates triangle inequality
  - $M_x = M_y = 0$: $D_{KL,\frac{\varepsilon_{(x|y)2}}{\varepsilon_{(x|y)0}}} \neq D_{KL,\frac{\varepsilon_{(x|y)0}}{\varepsilon_{(x|y)1}}}$; $D_{KL,\frac{\varepsilon_{(x|y)2}}{\tilde{\varepsilon}_{(x|y)0}}} \neq D_{KL,\frac{\varepsilon_{(x|y)0}}{\tilde{\varepsilon}_{(x|y)1}}}$. Not in line with reality.
- Non-fixed evaluation standards
  - $\boxed{\rho_1 = \rho_2 = 0}$: $D_f = 0$
  - $\boxed{\rho_1 \longrightarrow 0, \rho_2 \neq 0}$: $D_f = \lim_{\rho_1 \to 0} \rho_1 \ln(\rho_1) = 0$
  - $\boxed{\rho_2 \longrightarrow 0, \rho_1 \neq 0}$: $D_f \longrightarrow \infty$

    In this case, ρ_2 can be set to a small value c, but D_{KL} varies with different c, thus leading to different assessment standards.

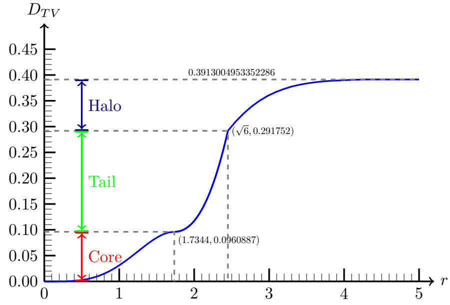
The Variation of D_f with Integration Radius

D_f in 4D spherical coordinates:

$$D_f[\rho_1'(r)||\rho_2'(r)] = \int \rho_2'(r) f\left[\frac{\rho_1'(r)}{\rho_2'(r)}\right] \cdot r^3 \sin^2\psi_1 \sin\psi_2 \, \mathrm{d}r\mathrm{d}\psi_1\mathrm{d}\psi_2\mathrm{d}\psi_3$$



The variation of $\rho'(r)$ with integration radius



The variation of D_{TV} with integration radius

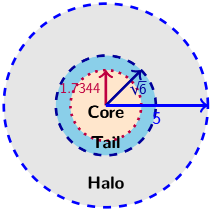
Different D_f Assign Different Weights to the Core, Tail and Halo

Distribution area	Gaussian-Water Bag			Gaussian-Parabolic		
	JS(%)	TV(%)	Hel(%)	JS(%)	TV(%)	Hel(%)
Core	14.04	24.56	33.99	14.86	27.86	34.48
Tail	34.49	50.00	29.34	27.28	51.97	23.65
Halo	51.47	25.44	36.67	57.86	20.17	41.87

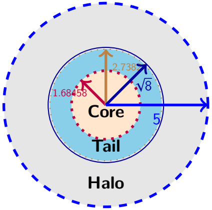
The sensitivity of various divergences to different beam distribution areas

Priority of divergence selection

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



Gaussian-Water Bag

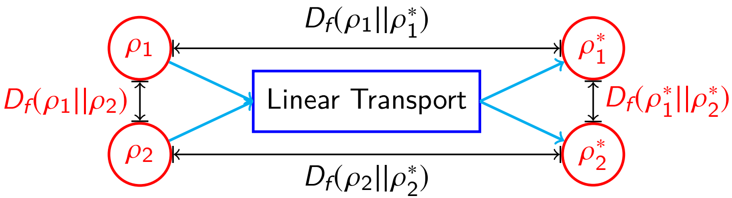


Gaussian-Parabolic

f-divergences in Linear Transport

Theorem

The f-divergences are conserved during the linear transport process.



- $D_f(\rho_1||\rho_1^*)$ not necessarily equal to $D_f(\rho_2||\rho_2^*)$;
- $D_f(\rho_1||\rho_2) = D_f(\rho_1^*||\rho_2^*)$ always holds true.

This conclusion may be utilized when describing beam transport using D_f.

Summary

1. Encouraging results from the quantification of beam distribution differences using common f-divergences
    - Addition tool to analyze beam simulations and experiments
2. f-divergence values from common 4D distributions can provide assessment standards
    - Only depend on mismatch factors and scaling ratios of the RMS emittances
3. Choice of f-divergences is goal-dependent
    - Different emphasis on core, tail and halo
4. Properties under transport have much to be explored
    - Conserved under linear transport

Thanks !