N-dimensional maximum-entropy tomography via particle sampling

Austin Hoover*

Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA

(Dated: March 10, 2025)

We propose a modified maximum-entropy (MENT) algorithm for six-dimensional phase space tomography. The algorithm uses particle sampling and low-dimensional density estimation to approximate large sets of high-dimensional integrals in the original MENT formulation. We implement this approach using Markov Chain Monte Carlo (MCMC) sampling techniques and demonstrate convergence of six-dimensional MENT on both synthetic and measured data.

I. INTRODUCTION

In plasma and beam physics, tomographic methods may be used to infer the distribution of particles in 2D, 4D, or 6D phase space from measured 1D or 2D projections. While 2D tomography is an established technique [1], 4D and 6D tomography introduce several challenges. The first challenge is to fit the data: searching the space of high-dimensional distribution functions is not straightforward [2–4]. The second challenge is to regularize the solution and quantify the reconstruction uncertainty: increasing the phase space dimension can generate an ill-posed inverse problem [5].

Entropy maximization [6, 7] is well-suited for high-dimensional tomography, which may generate a severely ill-posed inverse problem. The MENT algorithm [8–10] uses the method of Lagrange multipliers to generate an exact constrained entropy maximum via an unconstrained optimization. While MENT is valid for any number of dimensions and is compatible with arbitrary phase space transformations, its computational cost scales exponentially with the dimension of the phase space and the measurement resolution. While there have been several reported 4D MENT reconstructions [11–13], there have been no reported 6D reconstructions.

MENT struggles with high-dimensional problems because of the large number of high-dimensional integrals required to map the reconstructed distribution to its projections. We propose a modified algorithm that uses particle sampling and low-dimensional density estimation to approximate these integrals. We implement this approach using Markov Chain Monte Carlo (MCMC) sampling techniques and demonstrate convergence of 6D MENT for both synthetic and measured data sets.

II. BACKGROUND

Our task is to infer a probability distribution $\rho(\boldsymbol{x})$, defined over phase space coordinates $\boldsymbol{x} \in \mathbb{R}^N$, from its projections onto a set of M-dimensional planes. We assume the kth of K measurements occurs after a transformation $\mathcal{M}_k : \mathbb{R}^N \to \mathbb{R}^N$. We measure projections of the transformed coordinates

$$\boldsymbol{u}_k = \mathcal{M}_k(\boldsymbol{x}) \tag{1}$$

onto an axis $\boldsymbol{u}_{k_\parallel} \in \mathbb{R}^M$. The projection is the integral of $\rho(\boldsymbol{x})$ over the plane $\boldsymbol{u}_{k_\perp} \in \mathbb{R}^{N-M}$, which is orthogonal to $\boldsymbol{u}_{k_\parallel}$. The distribution must satisfy the following constraints:

$$G_k \left[\rho(\boldsymbol{x}) \right] = g_k(\boldsymbol{u}_{k_\parallel}) - \tilde{g}_k(\boldsymbol{u}_{k_\parallel}) = 0, \tag{2}$$

where $g_k(\boldsymbol{u}_{k_\parallel})$ are the *measured* projections and

$$\tilde{g}_k(\boldsymbol{u}_{k_\parallel}) = \int \rho(\boldsymbol{x}(\boldsymbol{u}_k)) d\boldsymbol{u}_{k_\perp} \tag{3}$$

are the *simulated* projections.

The constraints in Eq. (2) do not identify a unique distribution. To break the degeneracy, we introduce a prior distribution $\rho_*(\boldsymbol{x})$, which represents our confidence that a particle would land near $\boldsymbol{x}$ if sampled from the true distribution. We update the prior to a posterior $\rho(\boldsymbol{x})$ which satisfies the constraints *and* maximizes a functional $S[\rho(\boldsymbol{x}), \rho_*(\boldsymbol{x})]$, which we call the *entropy*. To determine the entropy functional, we invoke the principle of minimal updating: the prior should be updated as little as possible to fit the data. Enforcing this principle via specific axioms leads to a unique entropy [14]:

$$S[\rho(\boldsymbol{x}), \rho_*(\boldsymbol{x})] = - \int \rho(\boldsymbol{x}) \log \left(\frac{\rho(\boldsymbol{x})}{\rho_*(\boldsymbol{x})} \right) d\boldsymbol{x}. \tag{4}$$

The entropy is zero if and only if $\rho(\boldsymbol{x}) = \rho_*(\boldsymbol{x})$; otherwise it is negative.

The form of the posterior distribution can be derived from the functional

$$\Psi = S[\rho(\boldsymbol{x}), \rho_*(\boldsymbol{x})] + \sum_k \int \lambda_k(\boldsymbol{u}_{k_\parallel}) G_k \left[\rho(\boldsymbol{x}) \right] d\boldsymbol{u}_{k_\parallel}, \tag{5}$$

where $\lambda_k(\boldsymbol{u}_{k_\parallel})$ are Lagrange multiplier functions [10]. Enforcing zero variation of Ψ with respect to $\rho(\boldsymbol{x})$ and $\lambda_k(\boldsymbol{u}_{k_\parallel})$ gives

$$\begin{aligned}
\rho(\boldsymbol{x}) &= \rho_*(\boldsymbol{x}) \prod_k \exp\left(\lambda_k(\boldsymbol{u}_{k_\parallel}(\boldsymbol{x})) \right) \\
&= \rho_*(\boldsymbol{x}) \prod_k h_k(\boldsymbol{u}_{k_\parallel}(\boldsymbol{x})).
\end{aligned} \tag{6}$$

where we have defined the "component" functions $h_k(\boldsymbol{u}_k) = \exp(\lambda_k(\boldsymbol{u}_k))$.

If we find functions $h_k(\boldsymbol{u}_{k_\parallel})$ such that the distribution in (6) generates the measured projections, then we have found the constrained entropy maximum. We must solve the system of integral equations that results from substituting Eq. (6) into Eq. (2):

$$g_k(\boldsymbol{u}_{k_\parallel}) = \int \rho_*(\boldsymbol{x}) \prod_{k'} h_{k'}(\boldsymbol{u}_{k'_\parallel}) d\boldsymbol{u}_{k_\perp}. \qquad (7)$$

Note the hidden dependence of $\boldsymbol{x}$ and $\boldsymbol{u}_k$ on $\boldsymbol{u}_{k'}$ through Eq. (1). In practice, we evaluate the $h_k(\boldsymbol{u}_{k_\parallel})$ on a grid and define continuous functions by interpolation. In this case, the functions are defined by a set of positive numbers $\boldsymbol{\theta}$ (the function values on the grid) and we are left with an unconstrained optimization over $\boldsymbol{\theta}$.

III. METHODS

The MENT algorithm solves for the Lagrange multiplier functions using a Gauss-Seidel relaxation method

[8]. Note that $h_k(\boldsymbol{u}_{k_\parallel})$ is a function of $\boldsymbol{u}_{k_\parallel}$, which is orthogonal to the integration axis $\boldsymbol{u}_{k_\perp}$; thus, h_k can be factored out of the integral in Eq. (7). Rearranging gives a set of consistency conditions:

$$h_k(\boldsymbol{u}_{k_\parallel}) = \frac{g_k(\boldsymbol{u}_{k_\parallel})}{\int \rho_*(\boldsymbol{x}) \prod_{k'\neq k} h_{k'}(\boldsymbol{u}_{k'_\parallel}) d\boldsymbol{u}_{k_\perp}}. \qquad (8)$$

MENT begins by initializing the distribution to the prior within the measurement boundaries:

$$h_k^{(0)}(\boldsymbol{u}_{k_\parallel}) = \begin{cases} 1, & \text{if } g_k(\boldsymbol{u}_{k_\parallel}) > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (9)$$

where the superscript represents the iteration number. Next, for each measurement index k, h_k is set equal to the right-hand side of Eq. (8). The new value is immediately used to update h_{k+1}. In other words, for each iteration i and measurement index k,

$$h_k^{(i+1)}(\boldsymbol{u}_{k_\parallel}) = \frac{g_k(\boldsymbol{u}_{k_\parallel})}{\int \rho_*(\boldsymbol{x}) \left(\prod_{k'<k} h_{k'}^{(i+1)}(\boldsymbol{u}_{k'_\parallel}) \right) \left(\prod_{k'>k} h_{k'}^{(i)}(\boldsymbol{u}_{k'_\parallel}) \right) d\boldsymbol{u}_{k_\perp}}. \qquad (10)$$

The iterations may be written in a more interpretable form by dividing Eq. (10) by $h_k^{(i)}$:

$$\frac{h_k^{(i+1)}(\boldsymbol{u}_{k_\parallel})}{h_k^{(i)}(\boldsymbol{u}_{k_\parallel})} = \frac{g_k(\boldsymbol{u}_{k_\parallel})}{\int \rho_*(\boldsymbol{x}) \left(\prod_{k'<k} h_{k'}^{(i+1)}(\boldsymbol{u}_{k'_\parallel}) \right) \left(\prod_{k'\geq k} h_{k'}^{(i)}(\boldsymbol{u}_{k'_\parallel}) \right) d\boldsymbol{u}_{k_\perp}}. \qquad (11)$$

The $h_k^{(i)}$ term in the denominator was absorbed by the integral using Eq. (8). Comparing to Eq. (7), we see that the denominator in Eq. (11) is the kth *simulated projection* (Eq. (3)) of the MENT distribution function (Eq. (6)) using the up-to-date Lagrange multipliers. Denoting the simulated projection as $\tilde{g}_k(\boldsymbol{u}_{k_\parallel})$, Eq. (11) becomes

$$h_k^{(i+1)}(\boldsymbol{u}_{k_\parallel}) = h_k^{(i)}(\boldsymbol{u}_{k_\parallel}) \left[\frac{g_k(\boldsymbol{u}_{k_\parallel})}{\tilde{g}_k(\boldsymbol{u}_{k_\parallel})} \right], \qquad (12)$$

Finally, we add a learning rate, or damping coefficient, $0 < \omega \leq 1$, which slows the journey from the unconstrained to constrained maximum-entropy distribution to stabilize the iterations and handle noisy data [10]:

$$h_k^{(i+1)}(\boldsymbol{u}_{k_\parallel}) = h_k^{(i)}(\boldsymbol{u}_{k_\parallel}) \left[1 + \omega \left(\frac{g_k(\boldsymbol{u}_{k_\parallel})}{\tilde{g}_k(\boldsymbol{u}_{k_\parallel})} - 1 \right) \right]. \qquad (13)$$

Conventional MENT implementations have used numerical integration to simulate the projections $\tilde{g}_k(\boldsymbol{u}_{k_\parallel})$ of the distribution function $\rho(\boldsymbol{x})$ in Eq. (6). In this approach, the projected density is approximated by a sum

$$\tilde{g}_k(\boldsymbol{u}_{k_\parallel}) \approx \sum_i \rho(\boldsymbol{u}_{k_\perp}^{(i)} | \boldsymbol{u}_{k_\parallel}) \qquad (14)$$

over integration points $\{\boldsymbol{u}_{k_\perp}^{(i)}\}$. Let us assume that the integration points form a grid of size $R^{(N-M)}$ and that each projection is defined on a grid of size R^M, where R is the grid resolution. Then computing a single projection requires R^N evaluations of the probability density $\rho(\boldsymbol{x})$ defined in Eq. (6) (R^M total integrals, each of dimension $R^{(N-M)}$). These integrals can be prohibitively expensive for large phase space dimension N or measurement/integration grid resolution R.

We propose to estimate each projection from a batch

of particles $\{\boldsymbol{x}^{(i)}\}$ sampled from $\rho(\boldsymbol{x})$. There are a few potential advantages to this particle sampling approach. First, the phase space transformations are not inverted during the iterations and do not need to be symplectic. This is a convenience when interfacing with particle tracking codes. Second, the computational cost depends only weakly on the measurement resolution. Third, several techniques have been developed to sample from high-dimensional distribution functions without exponential storage requirements. Here, we focus on Markov Chain Monte Carlo (MCMC) sampling techniques [15], specifically the Metropolis-Hastings (MH) algorithm [16].

The MH algorithm explores the target distribution by a random walk, starting from point $\boldsymbol{x}_0$. Given the current point $\boldsymbol{x}_t$, a new point $\boldsymbol{x}_*$ is proposed from the "jumping" distribution $q(\boldsymbol{x}_*|\boldsymbol{x}_t)$. Given a random number $r \in [0, 1]$, $\boldsymbol{x}_t$ is updated as

$$\boldsymbol{x}_{t+1} = \begin{cases} \boldsymbol{x}_* & \text{if } r \leq \pi(\boldsymbol{x}_*|\boldsymbol{x}_t), \\ \boldsymbol{x}_t & \text{otherwise,} \end{cases} \qquad (15)$$

where

$$\pi(\boldsymbol{x}_*|\boldsymbol{x}_t) = \min\left(1, \frac{\rho(\boldsymbol{x}_*)}{\rho(\boldsymbol{x}_t)} \frac{q(\boldsymbol{x}_t|\boldsymbol{x}_*)}{q(\boldsymbol{x}_*|\boldsymbol{x}_t)}\right) \qquad (16)$$

is the acceptance probability. In the long run, the chain is guaranteed to converge to the target distribution $\rho(\boldsymbol{x})$. Although there are no guaranteed convergence rates, the algorithm finds wide use in practice. The performance depends primarily on the jumping distribution, which controls the acceptance probability. A common choice is a Gaussian jumping distribution centered on $\boldsymbol{x}_t$ with an adjustable covariance matrix. For common convergence diagnostics and tuning strategies, see [15].

Since MCMC is *gridless*, it has the potential to extend MENT to six or more dimensions. The problem is then to tune the sampling parameters such that each projection is a close approximation of the distribution function in Eq. (6). The distribution function may be expensive to evaluate because it requires K transformations of the coordinates through the particle tracking model. To increase the sampling speed, we therefore run multiple chains in parallel, evolving a small set of points through the phase space rather than a single point. We evolve the points until the chains "mix", i.e., until they are indistinguishable.

IV. DEMONSTRATIONS

The rest of this paper demonstrates particle-based 6D MENT in two examples. In the first example, we use MCMC sampling to reconstruct a 6D Gaussian Mixture (GM) distribution from its 2D marginals, showing that the MCMC approach is stable in at least one case with a known forward model and multimodal ground truth distribution. In the second example, we deploy the same algorithm to data from the Argonne Wakefield Accelerator (AWA).

Reconstructions are simplified if the scale and linear correlations between variables are known. The covariance matrix $\boldsymbol{\Sigma} = \langle \boldsymbol{x}\boldsymbol{x}^T \rangle$ provides this information. In the following examples, we assume the 2×2 block-diagonal elements of the covariance matrix are available from previous measurements and reconstruct the distribution in normalized coordinates $\boldsymbol{z} = \boldsymbol{V}^{-1}\boldsymbol{x}$, where $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{V}^T$ and $\boldsymbol{V}$ is a symplectic matrix [17].

Our first example uses a 6D Gaussian Mixture (GM) distribution as the ground truth. The distribution is the superposition of seven Gaussian distributions, each with a random mean and variance; it is a good candidate for initial testing because it is multimodal, easy to sample from, and has clear structure when projected onto lower dimensions. We collected the 2D marginal projections of the distribution as training data [18]. We used the Metropolis-Hastings (MH) algorithm to sample from the 6D MENT distribution defined by these projections. We used a Gaussian prior with covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is the assumed covariance matrix described above. We used a Gaussian jumping distribution with covariance matrix $0.25\hat{\boldsymbol{\Sigma}}$. We simulated the projections on each iteration by running 500 parallel MCMC chains for 1000 steps each to generate 5×10^5 total samples. The starting point of each chain was sampled from a Gaussian with covariance matrix $0.25\hat{\boldsymbol{\Sigma}}$. Gaussian filters were used to remove statistical noise from the binning process.

The MENT algorithm converged in 1-2 epochs, with each epoch taking less than 60 seconds. Fig. 1 shows the identical projections of the reconstructed and true distributions.

Our second example applies the same algorithm to the open-source data set and accelerator model described in [4]. The measurements include a set of x-y beam images after a quadrupole magnet, transverse deflecting cavity (TDC), and dipole bend at the Argonne Wakefield Accelerator (AWA). In [4], the Generative Phase Space Reconstruction (GPSR) algorithm was used to fit a 6D phase space distribution to 20 images, with each image corresponding to a different quadrupole, TDC, and dipole field strength. The accelerator was modeled in Bmad-X, a differentiable particle tracking code.

We assumed the block-diagonal elements of the covariance matrix were known from previous measurements [19]. We ran MENT for 5 iterations with learning rate $\omega = 0.25$, sampling 5×10^5 particles per iteration with 1000 parallel chains. With these parameters, each iteration took approximately 8 minutes on a 2021 MacBook Pro computer with an Apple M1 Max chip. (Tracking was performed on a CPU, not a GPU.)

Fig. 2 shows the measured and simulated images. Reasonable agreement is found after one or two iterations, with closer agreement after five iterations. Reconstructed features are similar to the GPSR reconstruction in [4], indicating the problem is well-constrained. Two-dimensional views of the reconstructed distributions are
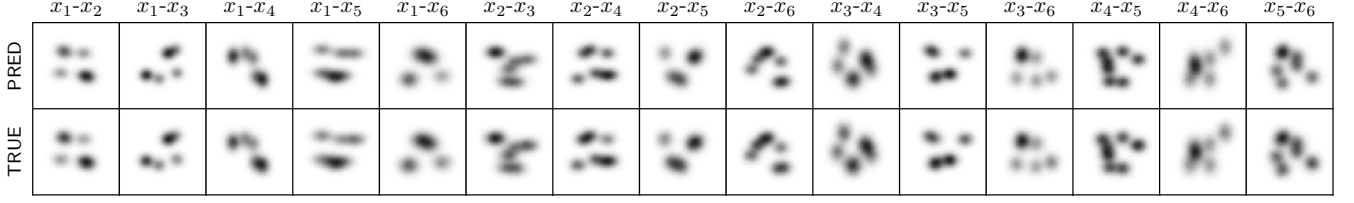
FIG. 1. Two-dimensional marginal projections of Gaussian Mixture distribution. Top: MENT prediction; bottom: ground truth.
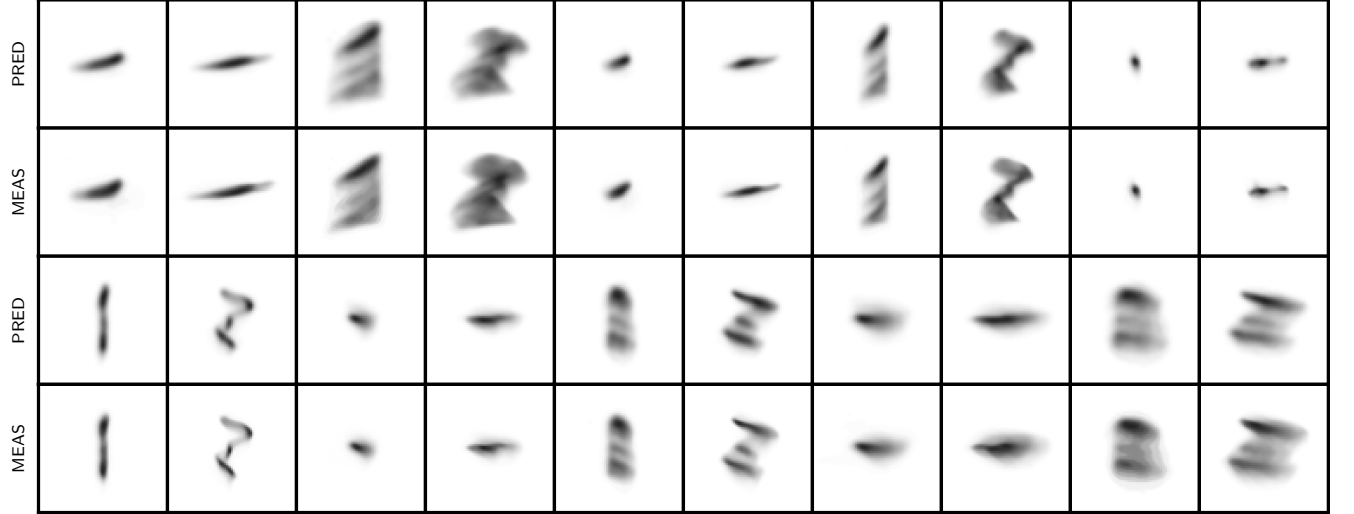

FIG. 2. 6D MENT applied to the Argonne Wakefield Accelerator (AWA) data set [4]. Rows 1/3 show simulated images and rows 2/4 show measured images.

included in the Supplemental Material. As a brief comparison, we sampled 9×10^4 particles from the MENT and GPSR models. For the training images, the average fractional error between measured and predicted pixels was 1.06% for GPSR and 0.82% for MENT; for the testing images, the error wass 0.93% for GPSR and 0.72% for MENT. We also estimated the entropy of each distribution by training a normalizing flow [20] on each set of 6D coordinates. The flow-based estimate gives an absolute entropy of 6.58 for GPSR and a slightly higher 7.61 for MENT. Although the MCMC sampling parameters could be tuned more methodically, we conclude that MENT has provided a plausible solution.

V. CONCLUSION

We have proposed a modified maximum-entropy (MENT) algorithm that scales to six-dimensional phase space. The method uses particle sampling, rather than numerical integration, to compute projections of the MENT distribution function. We showed that MCMC algorithms, such as the Metropolis-Hastings (MH) algorithm, can be used for six-dimensional reconstructions.

Particle-based MENT may be useful for accelerator physics research. MENT finds the most conservative distribution compatible with the data, making it a valuable tool when evaluating experimental results or comparing to other reconstruction algorithms. Interaction with trusted particle tracking codes is straightforward, as the only restriction on the tracking code is that particles must not interact with each other. Furthermore, there are no restrictions on the types of distributions MENT can represent. Possible applications and research topics include reconstructions in nonlinear lattices, reconstructions with self-consistent space charge forces, and high-dimensional and high-dynamic-range reconstructions [21].

Algorithmic improvements may involve tuning of the MCMC parameters [22] and investigation of alternative sampling methods such as dynamic nested sampling [23] or flow-based generative models [24]. Additionally, a major limitation of MENT is its lack of uncertainty quantification. MENT selects a single distribution $\rho(\boldsymbol{x})$— the simplest distribution allowed by the constraints—but does not consider the plausibility of other nearby distributions. Consider a vector $\boldsymbol{\theta}$ parameterizing the distribution $\rho(\boldsymbol{x}|\boldsymbol{\theta})$. The points $\{\boldsymbol{\theta}\}$ compatible with the

constraints form a *statistical manifold* [25]. We are ultimately interested in a probability distribution over $\boldsymbol{\theta}$, conditional on the measurements. Evaluating the shape of this higher-level distribution in the maximum-entropy framework is left as future work.

VI. ACKNOWLEDGEMENTS

VII. DATA AND CODE AVAILABILITY

The data and code used in this paper are available online [26].

---

[1] C. McKee, P. O'Shea, and J. Madey, Phase space tomography of relativistic electron beams, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **358**, 264 (1995).

[2] A. Wolski, D. C. Christie, B. L. Militsyn, D. J. Scott, and H. Kockelbergh, Transverse phase space characterization in an accelerator test facility, Phys. Rev. Accel. Beams **23**, 032804 (2020).

[3] A. Wolski, M. A. Johnson, M. King, B. L. Militsyn, and P. H. Williams, Transverse phase space tomography in an accelerator test facility using image compression and machine learning, Phys. Rev. Accel. Beams **25**, 122803 (2022).

[4] R. Roussel, J. P. Gonzalez-Aguilera, A. Edelen, E. Wisniewski, A. Ody, W. Liu, Y.-K. Kim, and J. Power, Efficient 6-dimensional phase space reconstruction from experimental measurements using generative machine learning (2024).

[5] A. Hoover and J. C. Wong, High-dimensional maximum-entropy phase space tomography using normalizing flows, Phys. Rev. Res. **6**, 033163 (2024).

[6] A. Caticha and R. Preuss, Maximum entropy and bayesian data analysis: Entropic prior distributions, Phys. Rev. E **70**, 046127 (2004).

[7] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Principles of maximum entropy and maximum caliber in statistical physics, Rev. Mod. Phys. **85**, 1115 (2013).

[8] G. Minerbo, MENT: A maximum entropy algorithm for reconstructing a source from projection data, Computer Graphics and Image Processing **10**, 48 (1979).

[9] N. J. Dusaussoy and I. E. Abdou, The extended MENT algorithm: a maximum entropy type algorithm using prior knowledge for computerized tomography, IEEE Transactions on Signal Processing **39**, 1164 (1991).

[10] C. Mottershead, Maximum entropy tomography, in *Proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods, Santa Fe, New Mexico, USA* (Springer, 1996) pp. 425–430.

[11] G. N. Minerbo, O. R. Sander, and R. A. Jameson, Four-dimensional beam tomography, IEEE Transactions on Nuclear Science **28**, 2231 (1981).

[12] J. C. Wong, A. Shishlo, A. Aleksandrov, Y. Liu, and C. Long, 4D transverse phase space tomography of an operational hydrogen ion beam via noninvasive 2d measurements using laser wires, Physical Review Accelerators and Beams **25**, 10.1103/PhysRevAccelBeams.25.042801 (2022).

[13] A. Hoover, Four-dimensional phase space tomography from one-dimensional measurements of a hadron beam, Phys. Rev. Accel. Beams **27**, 122802 (2024).

[14] This principle of minimal updating is enforced via two axioms which we paraphrase from [25]: (i) *Subset independence*: Probabilities conditioned on one domain should not be affected by information about a different, non-overlapping domain. Consequently, non-overlapping domains contribute additively to the entropy. (ii) *Subsystem independence*: If we assume two variables are unrelated, and if we receive information about each variable separately, then we should not infer any relationship between the variables on the basis of this information. In other words, if we measure the marginal distributions $\{\rho(\boldsymbol{v}), \rho(\boldsymbol{w})\}$, and if the prior does not encode any relationship between $\boldsymbol{v}$ and $\boldsymbol{w}$, then the posterior should be the product of the marginals: $\rho(\boldsymbol{v}, \boldsymbol{w}) = \rho(\boldsymbol{v})\rho(\boldsymbol{w})$.

[15] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis* (Chapman and Hall/CRC, 1995).

[16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics **21**, 1087 (1953).

[17] One way to fit the 4D or 6D covariance matrix to all measured projections is to leverage the Cholesky decomposition $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L}$ is a lower-diagonal matrix with positive diagonal elements, which ensures the covariance matrix is positive semi-definite. Alternatively, the diagonal or block-diagonal elements of the covariance matrix are typically available from faster measurements of the 2D phase space distributions $\{\rho(x, x'), \rho(y, y'), \rho(z, z')\}$. Cross-plane correlations ignored in this scenario are typically much weaker than within-plane correlations from, i.e., quadrupole focusing. If 2D measurements are unavailable, one may estimate the beam size by sampling from a uniform MENT prior and refining until the beam

size can be estimated. (The prior is nonzero only in the regions allowed by the measurements.).

[18] For a fixed number of measurements $K = N(N-1)/2$ of a distribution without linear correlations between planes, we hypothesize that the ideal set of measurements is the set of K two-dimensional marginals.

[19] In this case, we extracted the covariance matrix elements from the previous GPSR fit.

[20] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, Neural spline flows, Advances in neural information processing systems **32** (2019).

[21] A. Hoover, K. Ruisard, A. Aleksandrov, A. Zhukov, and S. Cousineau, Analysis of a hadron beam in five-dimensional phase space, Phys. Rev. Accel. Beams **26**, 064202 (2023).

[22] The distribution exhibits some fluctuations during training which may indicate that the sampling chains have not fully converged to the target distribution. However, the final distribution seems fairly stable with respect to random number generator seed. Additional convergence diagnostics (beyond visual checks) should be used in future studies to resolve this issue.

[23] E. Higson, W. Handley, M. Hobson, and A. Lasenby, Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation, Statistics and Computing **29**, 891 (2019).

[24] L. I. Midgley, V. Stimper, G. N. C. Simm, B. Schölkopf, and J. M. Hernández-Lobato, Flow annealed importance sampling bootstrap, in *NeurIPS 2022 AI for Science: Progress and Promises* (2022).

[25] A. Caticha, Entropy, information, and the updating of probabilities, Entropy **23**, 895 (2021).

[26] https://github.com/austin-hoover/ndim-ment-sampling.