

HIGH-DIMENSIONAL MAXIMUM-ENTROPY PHASE SPACE TOMOGRAPHY*

A. Hoover[†], Oak Ridge National Laboratory, Oak Ridge, TN, USA

Abstract

Reconstructing 4D or 6D phase space distributions from 1D or 2D measurements is a challenging inverse problem encountered in particle accelerators. Entropy maximization is an established method to incorporate prior information in the reconstruction, but it is typically infeasible in high-dimensional spaces. In this paper, I review two recent approaches to high-dimensional entropy maximization. The first approach utilizes differentiable simulations and a class of generative models known as *normalizing flows*, whereas the second approach employs the method of Lagrange multipliers and Markov Chain Monte Carlo (MCMC) sampling. My aim is to provide a short explanation of each method using a common notation. I conclude by mentioning several unsolved problems in phase space tomography.

INTRODUCTION

Phase space density measurements enable model-based prediction and control of charged particle beams in particle accelerators. This paper focuses on indirect measurements in which the density is *inferred* from partial information in the form of low-dimensional projections. Such measurements are especially challenging when the phase space is high-dimensional [1]. One challenge is to search the space of distribution functions. Simple models may not be able to capture the full complexity of the real distribution, while complicated representations may require too many parameters to fit to the data [2]. The second challenge is that the data do not identify a unique distribution; rather, they carve out a set of *feasible* distributions, each of which is compatible with the data to within the desired precision.

Here, I focus on the method of maximum entropy (MaxEnt) to select a distribution from the feasible set. After describing the reconstruction problem and MaxEnt strategy, I review two recent approaches to high-dimensional MaxEnt. The first approach leverages advances in differentiable simulations and generative modeling to find approximate maximum-entropy distributions. The second approach is an extension of the MENT algorithm, which uses the method of Lagrange Multipliers to find an exact constrained maximum entropy distribution. My aim is to provide a short explana-

tion of each method using a common notation, rather than to provide a comprehensive literature review. I conclude by mentioning several unsolved problems in phase space tomography, including the preservation of dynamic range, the inclusion of self-forces in the forward model, and robust uncertainty quantification (UQ). I do not discuss experimental design—a separate but equally important problem.¹

MAXIMUM-ENTROPY METHODS FOR PHASE SPACE RECONSTRUCTION

We begin by describing the generic reconstruction problem. Let $p(x)$ be the probability density at location $x \in \mathbb{R}^N$ in phase space, where $N \in \{2, 4, 6\}$ is the dimension of the phase space. Let s be the distance along the accelerator, with the reconstruction location at $s = 0$. Let ϕ be a vector of accelerator parameters, such as quadrupole strengths, which determines the single particle dynamics between reconstruction and measurement location. For measurement indices $k = 1, \dots, K$, we select accelerator parameters ϕ_k and tracking distance s_k . These choices define a set of symplectic transformations from input coordinates $x \equiv x_0$ to output coordinates x_k :

$$x_k = \mathcal{M}_k(x). \quad (1)$$

We measure the projection of the transformed distribution onto a plane $x_{k_\parallel} \in \mathbb{R}^M$, where $1 < M < N$. Here, x_{k_\parallel} identifies the observed coordinates on the measurement axis, while x_{k_\perp} identifies the unobserved coordinates. With these definitions, the projection is written as

$$g_k(x_{k_\parallel}) = \int p(\mathcal{M}_k^{-1}(x_k)) dx_{k_\perp}. \quad (2)$$

This setup is illustrated in Fig. 1. We wish to recover the initial distribution $p(x)$.

The data do not, in general, identify a unique distribution; the size of the solution space depends on the chosen coordinate transformations \mathcal{M}_k , the dimension of the phase space, and the dimension of the measurements. Rather than exploring the full solution space, let us instead search for a single distribution compatible with the data. Consider a *prior* distribution $p_*(x)$ over the phase space coordinates; how should we update this prior in light of new information? We will perform the update by maximizing an entropy functional $S[p(x), p_*(x)]$ subject to the measurement constraints:

$$\max_{p(x)} S[p(x), p_*(x)] \quad \text{s.t.} \quad G_k[p(x)] = 0, \quad (3)$$

where

$$G_k[p(x)] = g_k(x_{k_\parallel}) - \tilde{g}_k(x_{k_\parallel}) = 0 \quad (4)$$

¹ Implementations of the methods described here, along with example applications, are found in software repositories [3,4] and data repositories [5–7] linked to the corresponding papers [8–10].

* This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>)

[†] hooveram@ornl.gov

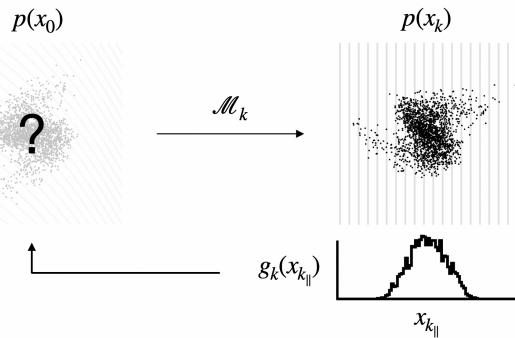


Figure 1: Illustration of the generic phase space reconstruction problem. We aim to recover a probability distribution from projections measured after a set of symplectic coordinate transformations. (Adapted from Ref. [9].)

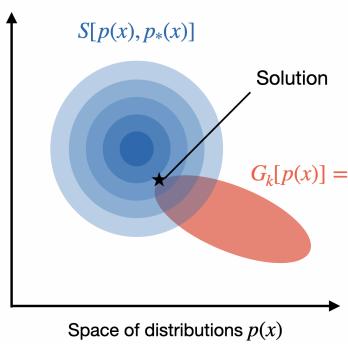


Figure 2: The method of maximum entropy (MaxEnt) selects the distribution of maximum relative entropy (blue) allowed by the measurement constraints (red). The solution is unique for the linear inverse problems described in this paper.

for measured projections g_k and simulated projections \tilde{g}_k . The entropy functional is fixed by a set of axioms² which enforce the following principle: *The chosen posterior distribution should coincide with the prior as closely as possible, and one should only update those aspects of one's beliefs for which corrective new evidence has been supplied* [11]. The resulting entropy is:

$$S[p(x), p_*(x)] = - \int p(x) \log \left(\frac{p(x)}{p_*(x)} \right) dx, \quad (5)$$

which lives in the range $[-\infty, 0]$ and reaches its maximum at the $p_*(x)$.

The optimization problem is illustrated in Fig. 2. The optimization is challenging both because of the difficulties in representing high-dimensional distributions and because of the hard constraints in Eq. (4). The following sections describe two recent approaches to the problem.

GENERATIVE MODELING APPROACHES

The first approach is based on the Generative Phase Space Reconstruction (GPSR) algorithm described in Ref. [10].

² See Refs. [11, 12] for details.

Generative models typically sample particles from a probability distribution by first sampling a random variable $z \in \mathbb{R}^{N'}$ from a simple *base distribution* $p(z)$, such as a Gaussian distribution, and then *unnormalizing* the samples:

$$z \sim p(z), \quad (6)$$

$$x = \mathcal{F}(z; \theta). \quad (7)$$

The unnormalized particles follow a distribution $p(x)$, even though the value of $p(x)$ may not be known. Thus, generative models provide easy-to-sample-from distributions of arbitrary complexity through the unnormalizing transformation \mathcal{F} . The unnormalizing transformation is typically implemented as an artificial neural network (NN) with parameters θ , enabling efficient estimation of gradients with respect to θ using automatic differentiation (AD) techniques [13].

In GPSR, the transfer maps \mathcal{M}_k in Eq. (1) are implemented using a differentiable physics simulation. This allows the propagation of gradients through the combined (unnormalization + lattice) transformation

$$x_k = \mathcal{M}_k(\mathcal{F}(z; \theta)). \quad (8)$$

The final step is the measurement operator, which is implemented by applying Kernel Density Estimation (KDE), a differentiable approximation to a histogram, to a batch of particles (projected onto the measurement axis). To simplify notation, introduce the function

$$\mathcal{P}_k(x) = x_{k||}, \quad (9)$$

which maps the phase space coordinates at $s = 0$ onto the k th measurement axis. Then, given normalized particle coordinates $\{z^{(i)}\}$, where i is the particle index, the projected density is estimated as

$$\begin{aligned} g_k(x_{k||}) &\approx \text{KDE}(\{\mathcal{P}_k(x^{(i)})\}) \\ &\approx \text{KDE}(\{\mathcal{P}_k(\mathcal{F}(z^{(i)}; \theta))\}). \end{aligned} \quad (10)$$

GPSR has been demonstrated in 4D and 6D phase space tomography, both in simulated and experimental reconstructions [10, 14].

To incorporate entropy in GPSR, we restrict \mathcal{F} to be a differentiable, invertible transformation called a *diffeomorphism*. The change in probability density under such a transformation is given by

$$\log p(x) = \log p(z) - \log |\det J_{\mathcal{F}}(z)|, \quad (11)$$

where $J_{\mathcal{F}} = dx/dz$ is the Jacobian matrix of the transformation. Generative models in which $\mathcal{F} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a diffeomorphism are known as *normalizing flows*, *flows*, or *flow-based models* [15]. By tracking the change in density in each layer of the transformation, flows provide access to the exact probability density in addition to the exact samples from the model. Such models can therefore estimate high-dimensional integrals in the form of expectation values. We are interested in the entropy integral of Eq. (5), which

may be estimated from n sampled particles $\{x^{(i)}\} \sim p(x)$ as follows:

$$\begin{aligned} S[p(x), p_*(x)] &= - \int p(x) \log(p(x)/p_*(x)) dx \\ &= -\mathbb{E}_{p(x)}[\log(p(x)/p_*(x))] \\ &\approx -\frac{1}{n} \sum_{i=1}^n \log(p(x^{(i)})/p_*(x^{(i)})). \end{aligned} \quad (12)$$

The estimate is simply the mean of $\log(p(x)/p_*(x))$ over the batch of particles.

In Ref. [16], Loaiza-Ganem, Gao, and Cunningham proposed the use of flows for entropy maximization. This technique was applied to the phase space reconstruction problem in Ref. [9]. It was found that an autoregressive flow architecture, with a masked neural network conditioner and rational-quadratic spline transformer, is a capable model with enough flexibility to represent complicated 6D distributions, but fast enough to train in a reasonable time. A simple penalty method was used with repeated minimization of the regularized loss function:

$$L(\theta) = -S(\theta) + \mu \sum_k D[g_k, \tilde{g}_k(\theta)], \quad (13)$$

where g_k is the measured projection, \tilde{g}_k is the simulated projection, and D is a discrepancy function such as the mean squared error. The loss is minimized for a series of penalty parameters $\{\mu_1 < \mu_2 < \dots\}$, starting from $\mu_1 = 0$, until the reduction in prediction error is negligible. Sample generation is approximately 50% slower than standard GPSR, and often around 5-10 penalty parameters are used, leading to typical training times of 5-20 minutes on a single GPU.

The studies in Ref. [9] showed that the entropy penalty has a positive effect, encouraging smooth distributions and tending toward known 2D maximum-entropy solutions. Figure 3 shows that the higher entropy distributions behave as expected for a Gaussian prior; for example, returning the product of the marginal distributions in the second column. 6D numerical experiments also show positive behavior compared to unregularized solutions, although comparison with exact maxent solutions is not possible.³

LAGRANGE MULTIPLIER METHODS

The traditional approach to entropy maximization is the Lagrange Multiplier Method [17]. This is the foundation of the MENT algorithm developed by Minerbo in 1971 [18]. We begin by constructing the Lagrangian Ψ :

$$\Psi = S[p(x), p_*(x)] + \sum_k \int G_k[p(x)] \lambda_k(x_{k_\parallel}) dx_{k_\parallel} \quad (14)$$

We then enforce zero change of Ψ under variations in λ_k and $p(x)$. The resulting expression for $p(x)$ is:

$$\begin{aligned} p(x) &= p_*(x) \prod_k \exp(\lambda_k(x_{k_\parallel})) \\ &= p_*(x) \prod_k h_k(x_{k_\parallel}). \end{aligned} \quad (15)$$

³ Cross-validation is one approach to regularizing the behavior of NN fits without an explicit regularization function. No cross-validation was used in these tests.

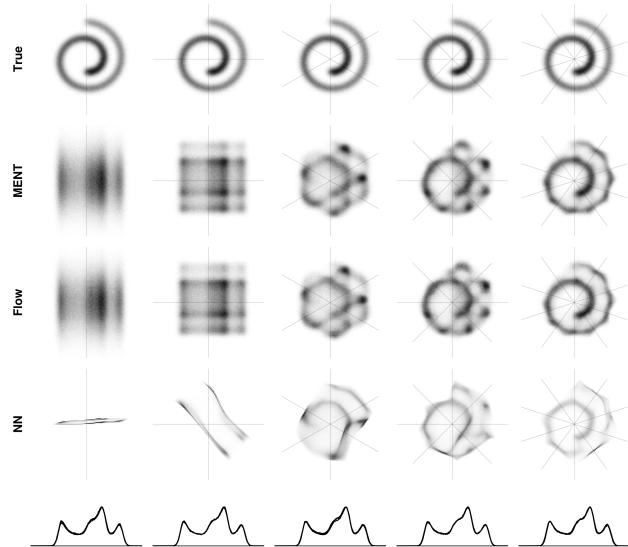


Figure 3: Entropy-regularized GPSR (MENT-Flow) can recover approximate maximum entropy distributions. The top row shows a ground-truth distribution, while the next three rows show MENT, MENT-Flow, and NN reconstructions with an increasing number of projections from left to right. The bottom row shows one of the projections in each case, with all four curves overlaid. (Adapted from Ref. [9].)

To make the dependence of x_k on $x = x_0$ explicit, we make use of the projection function \mathcal{P}_k in Eq. (9):

$$p(x) = p_*(x) \prod_k h_k(\mathcal{P}_k(x)). \quad (16)$$

Equation (16) gives the probability density at all points in phase space. Note that if we find Lagrange multiplier functions such that the distribution satisfies the constraints, we have found the *exact* maximum-entropy distribution.

Minerbo applied a nonlinear Gauss-Seidel method to solve for the Lagrange multiplier functions [18]. Plugging the distribution function, Eq. (16), into the constraint equations, Eq. (4), generates a simple algorithm that converges to the optimal distribution. The functions are initialized to constant values so that the distribution starts at the prior. Then the functions are updated as

$$h_k^{(i+1)} = h_k^{(i)} \left[(1 - \omega) + \omega \left(\frac{g_k}{\tilde{g}_k} \right) \right], \quad (17)$$

for damping rate $0 < \omega \leq 1$. Visual aids are provided in Figs. 4 and 5 shows an example maximum-entropy trajectory for a small value of ω . For typical values of ω , MENT converges in 1-2 iterations.

MENT was originally applied to a 4D reconstruction in 1981 [19]. This high-dimensional application of MENT was revived in Ref. [20] and continued in Ref. [21] at the Spallation Neutron Source (SNS) accelerator. Application to 6D tomography is more challenging. Each MENT iteration involves mapping the distribution function in Eq. (16)

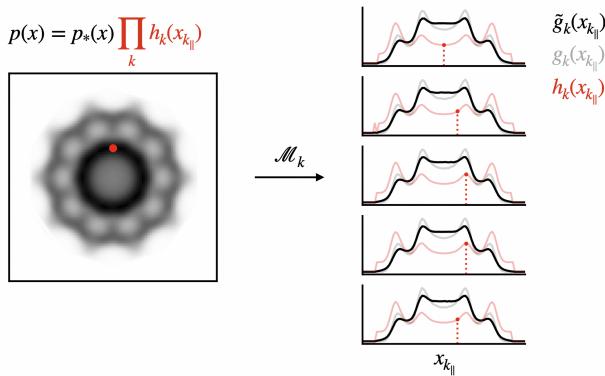


Figure 4: Illustration of the MENT algorithm. The density, Eq. (16), at a point is computed by mapping the point to each projection axis. The Lagrange multiplier functions h_k (red) are multiplied together, and the resulting value is multiplied by the prior. Each iteration, Eq. (17), uses the simulated projections \tilde{g}_k (black) and measured projections g_k (grey) to update the Lagrange multipliers.

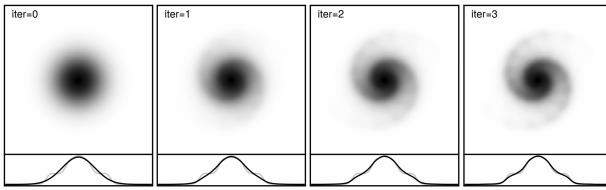


Figure 5: Maximum-entropy trajectory for small damping rate (ω). Structure is gradually added to the Gaussian prior.

to its projections. Conventional implementations use numerical integration to compute the integrals as shown in Fig. 6a. When the phase space is six-dimensional, the numerical integrals become four- or five-dimensional. The integrals are repeated for each point on the measurement axis; for example, at each pixel of a measured image. This numerical integration has a large computational cost making it infeasible for most 6D problems. For example, consider 15 projections of a 4D distribution onto a 50×50 measurement grid using a 50×50 integration grid. Each MENT iteration involves $15 \times 50 \times 50$ integrals. On a single CPU, each MENT iteration takes 60 seconds. When the distribution is 6D and the integration grid is $50 \times 50 \times 50 \times 50$, each MENT iteration takes **26 hours**.⁴

I propose to compute the projections by sampling particles from the distribution function, as illustrated in Fig. 6b. In Ref. [8], this approach was demonstrated with the Metropolis-Hastings (MH) algorithm, a simple variant of Markov Chain Monte Carlo (MCMC) sampling [22]. MH evolves particles through the phase space via a random walk such that, in the long run, the chain of particles converges to the target distribution. In Ref. [8], MH was used to demonstrate 6D MENT convergence on both a multimodal synthetic distribution as well as on experimental data, shared

in Ref. [10]. The runtime for the experimental 6D reconstruction, which involved 20 projections, was 7 minutes per iteration, converging in a few iterations.

UNSOLVED PROBLEMS

Uncertainty Quantification

The algorithms described here offer paths to maximum-entropy inference of high-dimensional phase space distributions. But even if these algorithms work perfectly, they do not provide a complete solution to the reconstruction problem. A complete solution would assign a probability to *all possible distributions*, then update these probabilities in response to data. We consider this a posterior distribution over the space of probability density functions. The maximum of the posterior, or MAP estimate, represents the most probable distribution (if there is a single maximum) and is thus a sensible choice if a single answer is desired. However, if the posterior is not sharply peaked, then there will be large uncertainty in this estimate.

In practice, we usually introduce a set of parameters θ to represent the distribution $p(x)$, writing either $p(x|\theta)$ or $p_\theta(x)$. We could also represent the measured projections using vectors η_k . Then, given a prior $P_*(\theta)$, we write the posterior probability of θ :

$$P(\theta) \propto P_*(\theta) \prod_k P(\eta_k|\theta). \quad (18)$$

The likelihood $\prod_k P(\eta_k|\theta)$ represents the probability of the data, given θ . With an error-free forward model and noiseless measurements, the likelihood is constant for distributions allowed by the constraints and zero otherwise. If we include random errors in the forward model, the likelihood will vary smoothly.

The prior $P_*(\theta)$ assigns a probability to each distribution logically prior to observing the data. Caticha [11] notes that the parameterized family of distributions $p(x|\theta)$ forms a *statistical manifold* with a unique metric known as the *Fisher information metric*. Caticha uses this metric to extend the maximum entropy technique beyond a point estimate and generate a full posterior over θ . It is unclear whether other choices of prior are justified based on simulated data or other assumptions.

If we can agree on a prior and likelihood, the next step is to characterize the posterior. A first step would be to quantify the posterior's width at its maximum, which provides a notion of uncertainty for the MAP estimate. This could be possible using a Gaussian approximation [23]. A next step would be to sample different θ vectors from the posterior. The samples represent an ensemble of distributions from which various statistics and visualizations could be produced. Sampling from the posterior seems to be a straightforward but possibly very high-dimensional sampling problem that could be attacked with a variety of modern algorithms.

Finally, it is important to include uncertainty in the accelerator lattice parameters ϕ , such as quadrupole strengths, which may not be known precisely [24]. In the Bayesian

⁴ https://github.com/austin-hoover/ment/blob/main/examples/rec_nd_2d/train_nd_marg_proj_reverse.ipynb.

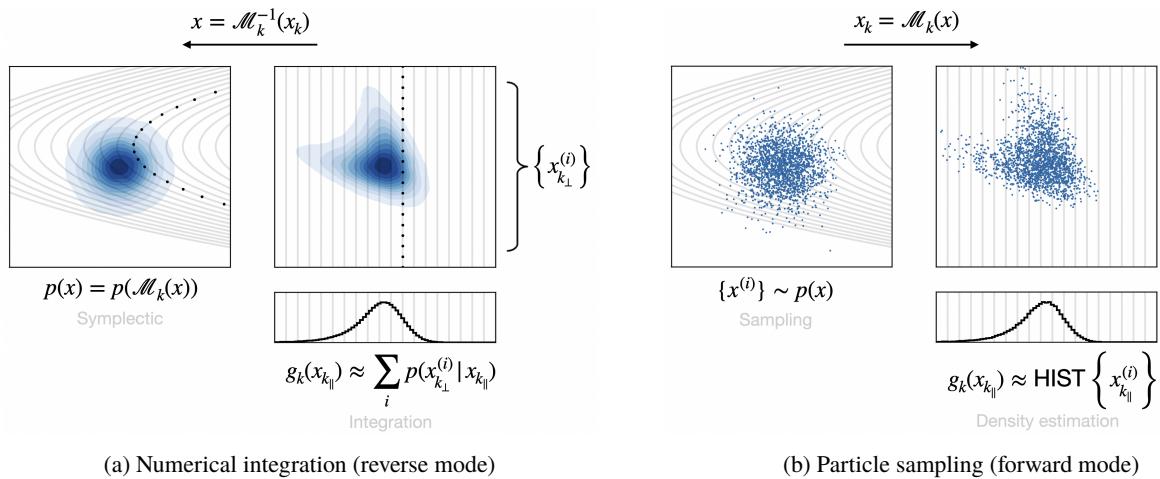


Figure 6: Two ways to implement the Gauss-Seidel iterations in the MENT algorithm. (a) Numerical integration: integration points are mapped backward to the reconstruction point to evaluate the probability density, then summed over to compute the projection. (b) Particle sampling: particles are sampled from the distribution and mapped forward to the observation point, and the projection is estimated using a histogram.

framework, this would require characterizing the joint posterior distribution $P(\theta, \phi)$ over both the lattice and beam parameters.

Dynamic Range

Low-density regions of phase space are important for predicting beam loss in high-power hadron accelerators [25]. Measurements sensitive to these low-density regions are considered *high-dynamic-range* (HDR), where the dynamic range is defined as the ratio of maximum to minimum measured density. Thus far, tomographic reconstructions have focused on low-dynamic-range (LDR) projections. GPSR is expected to struggle with HDR data, but MENT appears to be well-suited based on 2D tests. The limit on dynamic range in 4D/6D reconstructions is unknown.

Collective Effects

The inclusion of collective effects in the forward model dramatically complicates the reconstruction problem. In this case, the coordinate transformations in Eq. (1) depend on the initial distribution:

$$x_k = M_k(x, p(x)). \quad (19)$$

The resulting inverse problem is nonlinear, and most standard algorithms do not apply. On the other hand, it seems plausible that collective effects could place very strong constraints on the initial distribution, reducing the number of measurements required for high-dimensional reconstructions. The differentiable space charge model proposed in Ref. [26] could be used in a GPSR forward model, although it is unknown if this is feasible. It is unclear how to incorporate space charge into MENT.

SUMMARY

This article has reviewed two algorithms for maximum-entropy inference of phase space distributions from projection data.

GPSR is an efficient method to search the space of high-dimensional probability distributions if differentiable simulations are available. The entropic regularization discussed here could be improved by exploring new flow architectures to increase sampling speed, increase model flexibility, and decrease model size. In particular, it may be possible to use non-invertible unnormalizing transformations and directly compute the Jacobian matrix using automatic differentiation. The penalty method used to train the model could also be substituted for a more sophisticated constrained optimization algorithm with an automated stopping condition [27].

MENT is a surprisingly simple algorithm to fit high-dimensional distributions to measured projections. Its primary advantages are (i) exact constrained maximization, (ii) fast convergence, (iii) minimal storage requirements, (iv) no restrictions on the forward model, (v) the ability to represent every possible distribution function, and (vi) preservation of large dynamic range. In 2D and 4D tomography, MENT is an established method. The 6D (ND) version of the algorithm works but requires hand-tuning of the MCMC sampling algorithm. Future work will examine modern adaptive samplers [28] to more efficiently explore the phase space and more systematically tune the sampling parameters. It will also be important to parallelize the algorithm, which could provide significant speedups.

REFERENCES

- [1] S. Jaster-Merz *et al.*, “Experimental demonstration of a tomographic 5D phase-space reconstruction”, May 2025, arXiv:2505.13724 [physics].
doi:10.48550/arXiv.2505.13724

- [2] 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*, vol. 25, no. 12, p. 122 803, 2022.
doi:10.1103/PhysRevAccelBeams.25.122803
- [3] *Code repository: MENT*, <https://github.com/austin-hoover/ment>.
- [4] *Code repository: GPSR*, <https://github.com/roussel-ryan/gpsr>.
- [5] *Data repository: N-dimensional maximum-entropy tomography via particle sampling*.
doi:10.5281/zenodo.15413036
- [6] *Data repository: Four-dimensional phase space tomography from one-dimensional measurements of a hadron beam*.
doi:10.5281/zenodo.14807780
- [7] *Data repository: High-dimensional maximum-entropy phase space tomography using normalizing flows*.
doi:10.5281/zenodo.15528723
- [8] A. Hoover, “N-dimensional maximum-entropy tomography via particle sampling”, *Phys. Rev. Accel. Beams*, vol. 28, no. 8, p. L084601, 2025. doi:10.1103/zl2h-3v32
- [9] A. Hoover and J. C. Wong, “High-dimensional maximum-entropy phase space tomography using normalizing flows”, *Phys. Rev. Res.*, vol. 6, no. 3, p. 033 163, 2024.
doi:10.1103/PhysRevResearch.6.033163
- [10] R. Roussel *et al.*, “Efficient six-dimensional phase space reconstructions from experimental measurements using generative machine learning”, *Phys. Rev. Accel. Beams*, vol. 27, no. 9, p. 094 601, 2024.
doi:10.1103/PhysRevAccelBeams.27.094601
- [11] A. Caticha and R. Preuss, “Maximum entropy and Bayesian data analysis: Entropic prior distributions”, *Phys. Rev. E*, vol. 70, no. 4, p. 046 127, 2004.
doi:10.1103/PhysRevE.70.046127
- [12] A. Caticha, “Entropy, Information, and the Updating of Probabilities”, *Entropy*, vol. 23, no. 7, p. 895, 2021.
doi:10.3390/e23070895
- [13] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: A survey”, *J. Mach. Learn. Res.*, vol. 18, no. 153, pp. 1–43, 2018.
- [14] R. Roussel *et al.*, “Phase space reconstruction from accelerator beam measurements using neural networks and differentiable simulations”, *Phys. Rev. Lett.*, vol. 130, no. 14, p. 145 001, 2023.
doi:10.1103/PhysRevLett.130.145001
- [15] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, “Normalizing flows for probabilistic modeling and inference”, *J. Mach. Learn. Res.*, vol. 22, no. 57, pp. 1–64, 2021.
- [16] G. Loaiza-Ganem, Y. Gao, and J. P. Cunningham, “Maximum entropy flow networks”, Jan. 2017, arXiv:1701.03504 [stat.ME]. doi:10.48550/arXiv.1701.03504
- [17] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, MA, USA: Cambridge University Press, 2004.
- [18] G. Minerbo, “MENT: A maximum entropy algorithm for reconstructing a source from projection data”, *Comput. Graph. Image Process.*, vol. 10, no. 1, pp. 48–68, 1979.
doi:10.1016/0146-664X(79)90034-0
- [19] O. R. Sander, D. D. Chamberlin, R. A. Jameson, and G. N. Minerbo, “Beam tomography in two and four dimensions”, in *Proc. LINAC’79*, Montauk, NY, USA, Sep. 1979, pp. 314–318. <https://jacow.org/179/papers/S5-10.pdf>
- [20] 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”, *Phys. Rev. Accel. Beams*, vol. 25, no. 4, p. 042 801, 2022.
doi:10.1103/PhysRevAccelBeams.25.042801
- [21] A. Hoover, “Four-dimensional phase space tomography from one-dimensional measurements of a hadron beam”, *Phys. Rev. Accel. Beams*, vol. 27, no. 12, p. 122 802, 2024.
doi:10.1103/PhysRevAccelBeams.27.122802
- [22] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, Florida, USA: Chapman and Hall/CRC, 1995.
- [23] A. Krause and J. Hübotter, “Probabilistic artificial intelligence”, Feb. 2025, arXiv:2502.05244 [cs.AI].
doi:10.48550/arXiv.2502.05244
- [24] A. Wolski, D. Botelho, D. Dunning, and A. E. Pollard, “Accelerator beam phase space tomography using machine learning to account for variations in beamline components”, *J. Instrum.*, vol. 19, no. 07, p. P07013, 2024.
doi:10.1088/1748-0221/19/07/P07013
- [25] A. Aleksandrov, S. Cousineau, and K. Ruisard, “Understanding beam distributions in hadron linacs in the presence of space charge”, *J. Instrum.*, vol. 15, no. 07, p. P07025, 2020.
doi:10.1088/1748-0221/15/07/P07025
- [26] J. Qiang, “Differentiable self-consistent space-charge simulation for accelerator design”, *Phys. Rev. Accel. Beams*, vol. 26, no. 2, p. 024 601, 2023.
doi:10.1103/PhysRevAccelBeams.26.024601
- [27] S. Schechtman, D. Tiapkin, M. Muehlebach, and E. Moulines, “Orthogonal directions constrained gradient method: From non-linear equality constraints to Stiefel manifold”, in *Proc. COLT’2023*, Banhalore, India, pp. 1228–1258, Jul. 2023.
doi:10.48550/arXiv.2303.09261
- [28] N. Bou-Rabee, B. Carpenter, S. Liu, and S. Oberdörster, “The no-underrun sampler: A locally-adaptive, gradient-free mcmc method”, Jan. 2025, arXiv:2501.18548 [math.ST].
doi:10.48550/arXiv.2501.18548