The Earth Mover's Distance as a Measure of CP Violation

Tony Menzo

Department of Physics, University of Cincinnati, Cincinnati, OH, 45221, USA

in collaboration with Adam Davis, Ahmed Youssef, and Jure Zupan

Abstract: We introduce a new unbinned two sample test statistic sensitive to CP violation utilizing the optimal transport plan associated with the Wasserstein (earth mover's) distance. The efficacy of the test statistic is shown via one example of a CP asymmetric distribution: the Dalitz distributions of $B^0 \to K^-\pi^+\pi^0$ decays. The windowed version of the Wasserstein distance test statistic is shown to have comparable sensitivity to CP violation as the commonly used energy test statistic, but also retains information about the localized distributions of CP asymmetry over the Dalitz plot. For large statistic datasets we introduce two modified Wasserstein distance based test statistics - the binned and the sliced Wasserstein distance statistics, which show comparable sensitivity to CP violation, but improved time and space complexity scalings.

Introduction

CP violation within n-body weak decays (n>2) can manifest as local density asymmetries in phase space between conjugate decay probability distributions. These distributions may be analyzed on a so-called Dalitz plot as shown in Fig. (1).

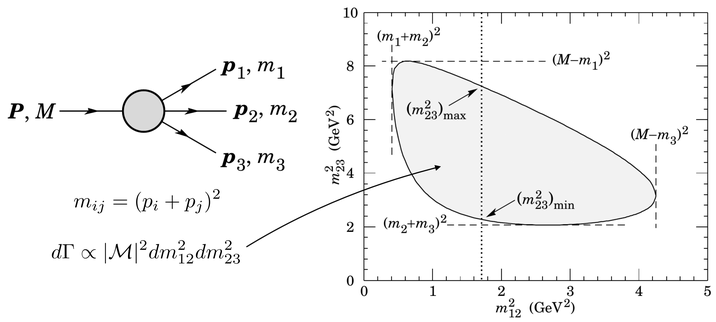


Figure 1: Left: In 3-body decays the differential decay rate is proportional to the matrix element squared. **Right:** An example of a 3-body decay Dalitz plot.

There are three ways to distinguish local CP violation on the Dalitz plot:

1. Model the conjugate decay amplitudes and perform a maximum likelihood fit on the model parameters noting that any significant difference between the conjugate model parameters is an indication of CP violation.

2. Perform a binned statistic analysis noting that significant bin count deviations between the conjugate distributions signals CP violation.

3. Perform an unbinned statistic analysis by comparing the nominal value of the test statistic with a model of the CP conserving distribution (the topic of this poster).

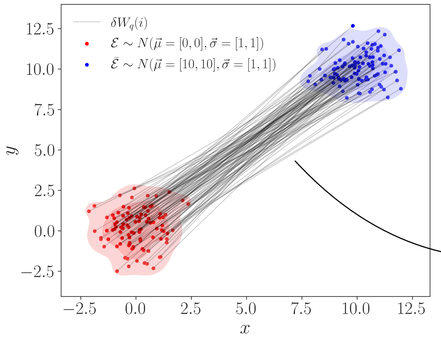
The Wasserstein distance or earth mover's distance (EMD)

The Wasserstein distance or earth mover's distance (EMD)[1] is a measure of similarity between two probability distributions. It is defined as the solution to the optimal transport problem which minimizes the work $d_{ij}f_{ij}$ (distance x mass) required to transport and reshape two probability distribution samples $\mathcal{E}$ and $\bar{\mathcal{E}}$ containing N and $\bar{N}$ events, respectively, into one another. This amounts to finding the optimal 'mass flow' matrix f_{ij} that determines the amount of weight to move from the i-th point in $\mathcal{E}$ and the j-th point in $\bar{\mathcal{E}}$.

$$W_q(\mathcal{E},\bar{\mathcal{E}}) = \left[\min_{\{f_{ij}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{\bar{N}} f_{ij}\left(\hat{d}_{ij}\right)^q\right]^{1/q}$$

$$\hat{d}_{ij} = \frac{1}{M^2}\sqrt{\sum_{i<j}^{3}(m_{ij}-\bar{m}_{ij})^2}, \qquad \sum_{i=1}^{N} f_{ij} = \frac{1}{\bar{N}}, \qquad \sum_{j=1}^{\bar{N}} f_{ij} = \frac{1}{N}, \qquad \sum_{i,j=1}^{N,\bar{N}} f_{ij} = 1$$

The value of the EMD can be visualized schematically as shown in Fig. (2) where the optimal transport configuration between two two-dimensional Gaussians is shown. The gray bands represent the (distance · weight) or $\delta W_q(i)$ between the i-th conjugate pair. The EMD is the sum of these contributions. See also Fig. (9).



$$W_q = \sum_{i}^{N}\delta W_q(i) = \sum_{\bar{i}}^{\bar{N}}\delta\bar{W}_q(\bar{i})$$

Figure 2: Example of a solution to the optimal transport problem between two-dimensional Gaussians. The sum of the distances δW_q chosen by the algorithm is approximately equal to the Wasserstein distance.

Similar distributions result in smaller values (~zero) of the EMD while dissimilar distributions result in larger values. **The EMD is thus sensitive to density asymmetries between samples and therefore well suited to be used as a test statistic that quantifies the amount of CP violation (CPV) in a physical system.**

Visualizing CP asymmetry on the Dalitz plot in $B^0 \to K^-\pi^+\pi^0$ decays using the Wasserstein distance

The optimal flow matrix f_{ij} in combination with the distance matrix d_{ij} may be used to visualize CP asymmetry across the Dalitz plot. We define two quantities within each bin of the Dalitz plot: the Wassersterin CP asymmetry as well as the direct CP asymmetry. The results for each asymmetry on CP violating (CPV) and CP conserving (CPC) datasets can be seen in Fig. (3).

$$\mathcal{W}^q_{\rm CP}(m_{12},m_{13}) = \frac{\sum_{\bar{i}}\delta\bar{W}_q(\bar{i}) - \sum_i \delta W_q(i)}{\sum_{\bar{i}}\delta\bar{W}_q(\bar{i}) + \sum_i \delta W_q(i)}, \qquad \mathcal{A}_{\rm CP}(m_{12},m_{13}) = \frac{d\bar{\Gamma}(\bar{m}_{12},\bar{m}_{13}) - d\Gamma(m_{12},m_{13})}{d\bar{\Gamma}(\bar{m}_{12},\bar{m}_{13}) + d\Gamma(m_{12},m_{13})}$$

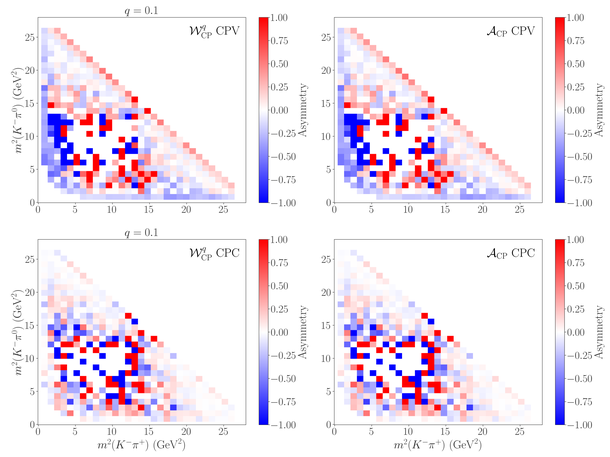


Figure 3: Binned Dalitz plot comparison between the Wasserstein asymmetry (left) and direct CP asymmetry (right), shown for CP violating $B^0 \to K^-\pi^+\pi^0$ decays (top) and CP conserving decays (bottom), i.e., decays in which the asymmetries in the amplitude model were set to zero. The results shown are normalized and averaged over 100 datasets, each containing 2N=2 x 10^3 (B and $\bar{B}$) events. **The agreement between the two asymmetries in each bin is within ~10%.**

As a global statistic

The Wasserstein test statistic can be used to assign a significance to the rejection of the CP conserving hypothesis. This is done by constucting a model of the CP conserving distribution and computing where the nominal value of the statistic resides on this distribution. The p-value is assigned according to the survival function i.e. 1-CDF as shown in Fig. (4).

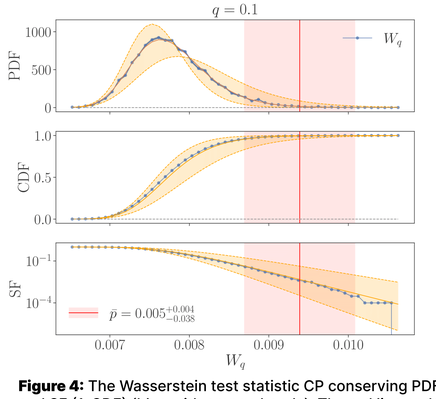


Figure 4: The Wasserstein test statistic CP conserving PDF, CDF, and SF (1-CDF) (blue with orange bands). The red line and band represent the average W_q +/- 1σ of 100 CPV data samples. **The test statistic nullifies the hypothesis of CPC at the ~3σ level.**

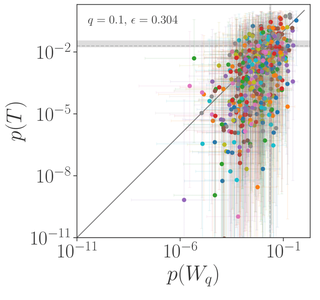


Figure 5: CPV sensitivity comparison between the Wasserstein test statistic W_q and energy test statistic T on like-datasets. The variable ϵ denotes the win percentage of W_q vs T (without taking into acount fit errors). The Wasserstein statistic is seen to be slightly less sensitive to CPV when compared to T.

Comparison to the Energy test

The energy test [2] is another unbinned two-sample test statistic commonly used in analyzing CP violation with phase space Dalitz data. The energy test statistic T utilizes distances between events plus a regulator function ψ containing a tuneable parameter σ that limits the sphere of influence around each event within the Dalitz plot. Its comparison in sensitivity to the Wasserstein test statistic can be seen in Fig. (5).

$$T = \sum_{i,j>i}^{N}\frac{\psi_{ij}}{N(N-1)} + \sum_{i,j>i}^{\bar{N}}\frac{\psi_{ij}}{\bar{N}(\bar{N}-1)} - \sum_{i,j}^{N,\bar{N}}\frac{\psi_{ij}}{N\bar{N}}, \qquad \psi_{ij} \equiv \psi(d_{ij};\sigma) = e^{-d_{ij}^2/2\sigma^2}$$

The windowed EMD

Because of the finite size of the datasets, the Wasserstein statistic is polluted by many small, non-zero contributions from CP conserving distances producing a long-tailed PDF and reducing sensitivity to CPV. To mitigate this we introduce the windowed Wasserstein statistic, I_q, which limits the distance contributions to specified regions of δW_q as shown in Figs. (6), (7), (8).

$$I_q \equiv \sum_i w\left(\delta W^{\rm win}_{\rm min}, \delta W^{\rm win}_{\rm max}, \delta\overline{W^{\rm win}_{\rm min}}, \delta\overline{W^{\rm win}_{\rm max}}; \delta W_i\right), \qquad w(x) = \begin{cases} +1 & x \in [\delta W^{\rm win}_{\rm min}, \delta W^{\rm win}_{\rm max}], \\ -1 & x \in [\delta\overline{W^{\rm win}_{\rm min}}, \delta\overline{W^{\rm win}_{\rm max}}], \\ 0 & \text{otherwise.} \end{cases}$$

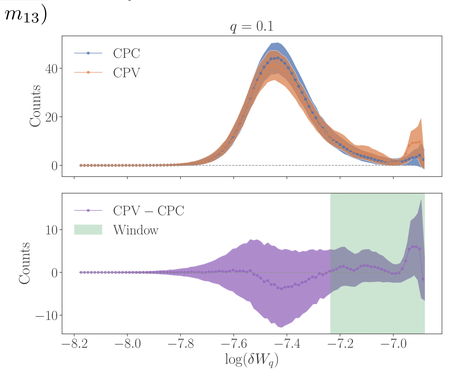


Figure 6: Top: A comparison of histograms between $\log(\delta W_q)$ values of CPC and CPV datasets. **Bottom:** The difference in CPC and CPV histograms as well as the selected window region (chosen considering the surplus in counts as well as the +/- 1σ dataset ensemble average (bands)).

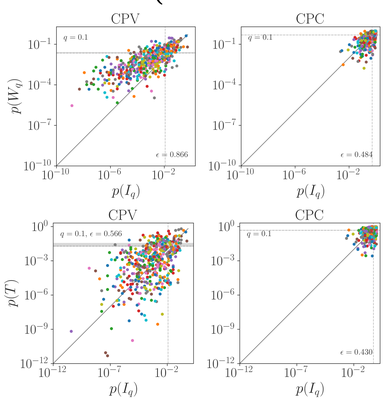


Figure 7: CPV sensitivity comparison between the windowed statistic I_q with W_q and T on CPV (left) and CPC (right) samples. **We see that the windowed statistic is comparable/slightly more sensitive to CPV than the energy test statistic.**

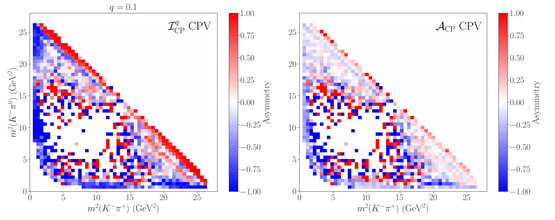


Figure 8: To visualize where the selected window-regions of CPV are on the Dalitz plot, we introduce the windowed equivalent of the Wasserstein CP asymmetry parameters within in each bin. We see that the chosen window in Fig. (6) is correctly filtering CPC distances and retaining CPV distances.

The binned and sliced EMD statistics

Solving the earth mover's distance transport problem optimally scales as ~O(N^3) in time complexity and O(N^2) in space complexity. Thus, for large datasets containing N~10^6 decay events the Wasserstein distance test statistic is computationally prohibitive. To remedy this we introduce two modified version of the statistic with improved time and space complexity scalings at ~ no loss in sensitivity to CPV (for small datasets). The first is the binned Wasserstein statistic which swtiches from encoding the CPV in distances between datasets to an over/under abundance of weights over localized bins between datasets as shown in Fig. (9). This scales as ~O(n^3) in time and ~O(n^2) in space complexities where n is the total number of bins used in the binned and projected Dalitz plot. The second variant uses the sliced Wasserstein distance which approximates the original W_q as the average of $N_{\rm slices}$ one-dimensional W_q projections. The sliced Wasserstein distance scales as ~O(N log(N)) in time and ~O(N) in space complexities. Of course, one could combine these and create a binned-sliced W_q with further improvements to the time and space complexity scalings (~O(n log(n)) time, ~O(n) space).

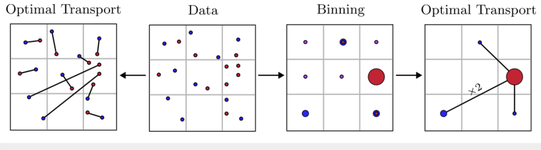


Figure 9: Pictorial comparison between unbinned (left) and binned (right) Wasserstein statistic methods. Note how in the binned case, the optimal transport algorithm effectively sets to zero in the last step the number counts in the bins that have the same counts between the two CP conjugate datasets (red and blue).

References/Acknowledgments

[1] Panaretos, Victor M., and Yoav Zemel. "Statistical aspects of Wasserstein distances." Annual review of statistics and its application 6 (2019): 405-431.

[2] Williams, Mike. "Observing CP violation in many-body decays." Physical Review D 84.5 (2011): 054015.

TM acknowledges support in part by the DOE grant de-sc0011784 and NSF OAC-2103889.

Public code: https://github.com/adamdddave/EMD4CPV