FERMILAB-PUB-14-396-T, LPT-Orsay-14-76, nuhep-th/14-06

arXiv:1410.4817v1 [hep-ph] 17 Oct 2014

Technical Note for 8D Likelihood Effective Higgs Couplings Extraction Framework in $h \to 4\ell$

Part I: From Generator Level to Detector Level

Yi Chen [a], **Emanuele Di Marco** [a], **Joe Lykken** [b], **Maria Spiropulu** [a], **Roberto Vega-Morales** [b,c,d], **Si Xie** [a]

[a]Lauritsen Laboratory for High Energy Physics, California Institute of Technology, Pasadena, CA, 92115, USA
[b]Theoretical Physics Department, Fermilab, P.O. Box 500, Batavia, IL 60510, USA
[c]Laboratoire de Physique Théorique, CNRS - UMR 8627, Université Paris-Sud, Orsay, France
[d]Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA

Abstract

In this technical note we present technical details on various aspects of the framework introduced in [1] aimed at extracting effective Higgs couplings in the $h \to 4\ell$ 'golden channel'. Since it is the primary feature of the framework, we focus in particular on the convolution integral which takes us from 'truth' level to 'detector' level and the numerical and analytic techniques used to obtain it. We also briefly discuss other aspects of the framework.

1 Introduction

It is well known that the $h \to 4\ell$ ($4\ell = 2e2\mu, 4e, 4\mu$) 'golden channel' is a powerful means of studying the Higgs couplings to neutral electroweak gauge bosons and various methods have long been proposed for studying it [2–6] and more recently [7–36]. Though 'truth' level (or generator) studies of the golden channel give a good approximate estimate of the expected sensitivity to the Higgs ZZ, $Z\gamma$, and $\gamma\gamma$ couplings [36], when analyzing data obtained at the LHC (or future colliders) a detector level likelihood which accounts for the various detector effects is necessary. Since generally detector level likelihoods are obtained via the use of Monte Carlo methods, it becomes difficult to obtain the full multi-dimensional likelihood for the 4ℓ final state. Typically one needs to fill large multi-dimensional templates that require an impractical amount of computing time. There are also potential collateral binning and 'smoothing' side-effects often associated with these methods. In the case of the golden channel this necessitates the use of kinematic discriminants which 'collapse' the fully multi-dimensional likelihood into two or perhaps three detector level observables [31]. This approach is normally taken to facilitate the inclusion of detector effects, but is not optimal when fitting to a large number of parameters simultaneously [16, 37]. This is unfortunate in the case of the golden channel where in principle there are twelve observables which can be used to extract a large number of parameters at once, including their correlations. It would be satisfying and useful to have a framework which is free of these issues and capable of utilizing all available information in the four lepton final state at detector level.

This is accomplished in our framework [1] by performing an explicit convolution of the generator ('truth') level probability density, formed out of analytic expressions for the signal and

Operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy.

background differential cross sections, with a transfer function which encapsulates the relevant detector effects. This can be represented schematically as follows,

$$P(\vec{X}^{\rm R}|\vec{\mathcal{A}}) = \int P(\vec{X}^G|\vec{\mathcal{A}})T(\vec{X}^R|\vec{X}^G)d\vec{X}^G. \tag{1}$$

Here we take $\vec{X}$ to represent the full set of center of mass variables, of which there are twelve in the golden channel, to be discussed more below, and $\vec{\mathcal{A}}$ represents some set of lagrangian parameters [1]. The transfer function $T(\vec{X}^R|\vec{X}^G)$ takes us from generator (G) level to reconstructed (R) (or detector) level observables and represents the probability of reconstructing the observables $\vec{X}^R$ given the generator level observable $\vec{X}^G$. It is treated as a function of $\vec{X}^R$ which takes $\vec{X}^G$ as input. As will be described more below, once the integration in Eq.(1) is performed we must then normalize over all twelve reconstructed level observables to obtain the detector level probability density function (*pdf*). After performing this 12-dimensional integration and normalizing, we are left with a *pdf* from which *we construct an un-binned twelve-dimensional detector level likelihood* which is a continuous function of the effective couplings (or Lagrangian parameters) and takes as its input, up to twelve reconstructed, detector-level center of mass observables. In the current implementation [1] we average over the four production variables to reduce the systematic uncertainties, thus obtaining an eight-dimensional likelihood in terms of just decay observables. However, this step is in principle not necessary.

We have performed the integration in Eq.(1) for both the $h \to 4\ell$ signal as well as the dominant $q\bar{q} \to 4\ell$ background (computed in [18, 34, 38]) and emphasize that it has not been done via Monte Carlo methods, but instead by a combination of numerical and analytical techniques to be discussed in detail below. With these detector level *pdfs* in hand we can go on to perform fast and accurate multi-parameter extractions on data obtained at colliders in the 4ℓ channel as demonstrated in [1] as well as in recent CMS studies [39]. In this technical note we present details on various aspects of how the convolution integral is performed and briefly discuss other aspects of the framework. Further details can be found in accompanying studies [1,18,34,36,38,40] as well as [39].

2 From 'Truth' Level to 'Detector' Level

We now describe in more detail how the convolution integral in Eq.(1) is performed. We first review the twelve center of mass variables present in the 4ℓ system before describing the various numerical and analytic techniques used in the integration which takes us from 'truth' level to detector level.

2.1 Center of Mass Observables

The twelve variables are discussed in more detail in [1, 17, 18, 34, 36], but are listed here for convenience as they will be used extensively in what follows. The invariant masses are defined as:

- $\sqrt{\hat{s}} \equiv M_{4\ell} \equiv m_h$ – The invariant mass of the four lepton system or the Higgs mass in case of signal.
- M_1 – The invariant mass of the lepton pair system which reconstructs closest to the Z mass.
- M_2 – The invariant mass of the other lepton pair system and interpreted as $M_2 < M_1$. This condition holds as long as $\sqrt{\hat{s}} \lesssim 2M_Z$.

These invariant masses are all independent subject to the constraint $(M_1 + M_2) \leq \sqrt{\hat{s}}$ and serve as the most strongly discriminating observables between different signal hypothesis as well as between signal and background. Note also that the $4e/4\mu$ final state can be reconstructed in two different ways due to the identical final state interference. This is a quantum mechanical effect that occurs at the amplitude level and thus both reconstructions are valid. The definitions M_1 and M_2 remain unchanged however. The angular variables are defined as:

- Θ – The production angle between the momentum vectors of the lepton pair which reconstructs to M_1 and the total 4ℓ system momentum.
- $\theta_{1,2}$ – Polar angle of the momentum vectors of e^-, μ^- in the lepton pair rest frame.
- Φ_1 – The angle between the plane formed by the M_1 lepton pair and the 'production plane' formed out of the momenta of the incoming partons and the momenta of the two lepton pair systems.
- Φ – The angle between the decay planes of the final state lepton pairs in the rest frame of the 4ℓ system.

We group the angular variables as follows $\vec{\Omega} = (\Theta, \cos\theta_1, \cos\theta_2, \Phi_1, \Phi)$. These angular variables are useful in aiding to distinguish different signal hypothesis and in particular between those with different CP properties, as well as in discriminating signal from background. Lastly, we have production variables associated with the initial partonic state four momentum:

- $\vec{p}_T \equiv (p_T \cos\phi_{4\ell}, p_T \sin\phi_{4\ell})$ – The momentum in the transverse direction of the 4ℓ system.
- Y – Defined as the motion of the 4ℓ system along the longitudinal direction.
- ϕ – Defines a global rotation of the event in the 4ℓ rest frame.

2.2 Changing Variables for Background *pdf*

Beginning from Eq.(1) we first discuss the construction of the background detector level *pdf*. The construction of the signal will be discussed separately as there is a subtle, but important, difference in performing the convolution. Since there are no undetermined parameters in the background the generator and detector-level (un-normalized) differential cross sections are given simply by $P_B(\vec{X}^G)$ and $P_B(\vec{X}^R)$ respectively and the convolution can be written as,

$$P_B(\vec{X}^{\mathrm{R}}) = \int P_B(\vec{X}^G) T(\vec{X}^R|\vec{X}^G) d\vec{X}^G. \tag{2}$$

The set of variables $\vec{X} \equiv (\vec{p}_T, Y, \phi, \hat{s}, M_1, M_2, \vec{\Omega})$ exhausts the twelve degrees of freedom (note that $\vec{p}_T$ has 2 components and $\vec{\Omega}$ contains 5 angles) available to the four (massless) final state leptons. The differential volume element is given by $d\vec{X} = d\hat{s} dM_1^2 dM_2^2 d\vec{\Omega} \cdot d\vec{p}_T dY d\phi$.

To perform this convolution with the transfer function we must first transform to the basis in which the detector smearing of the lepton momenta is parameterized. This requires that we first transform from the basis of the twelve center of mass variables defined in Sec. 2.1 to the three momentum basis for the four final state leptons. This can be represented as follows,

$$\begin{aligned} P_B(\vec{X}^{\mathrm{R}}) &= \int P_B(\vec{X}^G) T(\vec{X}^R|\vec{X}^G) d\vec{X}^G \\ &= \int P_B(\vec{X}^G) T(\vec{P}^R|\vec{P}^G) \frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|} d\vec{P}^G, \end{aligned} \tag{3}$$

where the differential volume element is now given by,

$$d\vec{P}^G = \prod_{i=1}^{4} d\vec{p}_i^{\,G}, \tag{4}$$

and $\vec{p}_i^{\,G}$ is the generator level three momentum of the i'th lepton. The $|\mathbf{J}_G^{\vec{P}}|$ is the Jacobian associated with the twelve dimensional change of variables from $\vec{X}^G \to \vec{P}^G$ in the differential volume element. The $|\mathbf{J}_R^{\vec{P}}|$ arises from the change of variables $\vec{X}^R \to \vec{P}^R$ in the transfer function (remembering $T(\vec{X}^R|\vec{X}^G)$ is treated as a function of $\vec{X}^R$) which we loosely also refer to as a Jacobian, as we will do for all subsequent change of variables to follow. Ideally to find these Jacobian factors one should construct the 12×12 matrix associated with these transformations and then calculate the determinant. However, since these transformations are highly non-linear and must be performed for each point in phase space, this is untenable analytically. We therefore implement a numerical algorithm to calculate these factors for each phase space point which will be discussed in more detail in Sec. 2.4.

Since we make the assumption that detector smearing will only affect the component of the lepton momentum parallel to the direction ($p_{i||}$) of motion and not the two components perpendicular to the direction of motion ($\vec{p}_{i\perp}$) (which are zero at generator level) we find it convenient to decompose the lepton three momenta $\vec{p}_i$ in terms of $p_{i||}$ and $\vec{p}_{i\perp}$. Note that this assumption is equivalent to assuming angular resolution effects due to detector smearing can be neglected, which is an excellent approximation for the LHC detectors [41,42]. In the $(p_{i||}, \vec{p}_{i\perp})$ basis only the transfer function associated with $p_{i||}$ is non-trivial while the one associated with the perpendicular components can be represented simply as a delta function for each perpendicular direction, thus allowing for trivial integration over the eight $\vec{p}_{i\perp}$ variables.

The differential volume element can now be written as,

$$d\vec{P}^G = \prod_{i=1}^{4} d\vec{p}_i^{\,G} = \prod_{i=1}^{4} d\vec{p}_{i\perp}^{\,G} dp_{i||}^{\,G}. \tag{5}$$

We then use the property of the transfer function that it is explicitly parametrized in terms of the ratio of reconstructed and generator level momentum components along the direction of motion to again change variables as follows,

$$\begin{aligned} P_B(\vec{X}^{\mathrm{R}}) &= \int P_B(\vec{X}^G) T(\vec{P}^R|\vec{P}^G) \frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|} \prod_{i=1}^{4} d\vec{p}_{i\perp}^{\,G} dp_{i||}^{\,G} \\ &= \int P_B(\vec{X}^G) T(\vec{c}\,|\vec{P}^G) \frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|} \frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|} \prod_{i=1}^{4} dc_i d\vec{p}_{i\perp}^{\,G}, \end{aligned} \tag{6}$$

where we have defined $c_i = p_{i||}^R / p_{i||}^G$ and $\vec{c} = (c_1, c_2, c_3, c_4)$. The components of the Jacobians $|\mathbf{J}_R^{\vec{c}}|$ and $|\mathbf{J}_G^{\vec{c}}|$ which take us from $p_{i||}^R \to c_i$ and $p_{i||}^G \to c_i$ variables are obtained trivially as $|1/p_{i||}^G|$ and $|c_i/p_{i||}^G|$ for the transfer function (which is now a function of $\vec{c}$) and the differential volume element respectively. Finally, we use the fact that $c_1 c_2 = (M_1^R/M_1^G)^2$ and $c_3 c_4 = (M_2^R/M_2^G)^2$ to eliminate c_2 and c_4 and make a final change of variables to the basis in which we perform the explicit four

dimensional integration,

$$
\begin{aligned}
P_B(\vec{X}^{\rm R}) &= \int P_B(\vec{X}^G) T(\vec{c}\,|\vec{P}^G) \frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|} \frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|} \prod_{i=1}^{4} dc_i \\
&= \int P_B(\vec{X}^G) T(\vec{c}\,|\vec{P}^G) \times \frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|} \frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|} |\mathbf{J}_B^{\vec{M}}| \cdot dc_1 dc_3 dM_1^{2G} dM_2^{2G},
\end{aligned} \tag{7}
$$

where in the first line in Eq.(7) we have implicitly used the delta functions in the transfer function to perform the eight dimensional integration over $\vec{p}_{i\perp}^{\,G}$. The Jacobian $|\mathbf{J}_B^{\vec{M}}|$ is obtained analytically from the change of variables $c_2, c_4 \to M_1^{G^2} M_2^{G^2}$ by observing that,

$$
\begin{aligned}
(p_1^G + p_2^G)^2 &= M_1^{G^2}, \\
(p_3^G + p_4^G)^2 &= M_2^{G^2}, \\
(p_1^R + p_2^R)^2 &= (c_1 p_1^G + c_2 p_2^G)^2 = M_1^{R^2}, \\
(p_3^R + p_4^R)^2 &= (c_3 p_3^G + c_4 p_4^G)^2 = M_2^{R^2}.
\end{aligned} \tag{8}
$$

Assuming that the leptons are massless (an excellent approximation for muons and electrons) and expanding out the equations we arrive at,

$$
\begin{aligned}
M_1^{R^2} = 2c_1 c_2 p_1^G p_2^G = c_1 c_2 M_1^{G^2}, \\
M_2^{R^2} = 2c_3 c_4 p_3^G p_4^G = c_3 c_4 M_2^{G^2},
\end{aligned} \tag{9}
$$

From here we solve for the smearing factors c_2 and c_4,

$$
\begin{aligned}
c_2 &= \frac{1}{c_1} \frac{M_1^{R^2}}{M_1^{G^2}} \equiv \frac{1}{c_1} R_{12} \\
c_4 &= \frac{1}{c_3} \frac{M_2^{R^2}}{M_2^{G^2}} \equiv \frac{1}{c_3} R_{34}.
\end{aligned} \tag{10}
$$

from which the Jacobian elements are easily computed as,

$$
\begin{aligned}
dc_2 = -\frac{1}{c_1} \frac{M_1^{R^2}}{M_1^{G^4}} dM_1^{G^2} \\
dc_4 = -\frac{1}{c_3} \frac{M_2^{R^2}}{M_2^{G^4}} dM_2^{G^2}.
\end{aligned} \tag{11}
$$

This gives finally for $|\mathbf{J}_B^{\vec{M}}|$,

$$
|\mathbf{J}_B^{\vec{M}}| = \frac{1}{c_1} \frac{M_1^{R^2}}{M_1^{G^4}} \frac{1}{c_3} \frac{M_2^{R^2}}{M_2^{G^4}}. \tag{12}
$$

We thus see in Eq.(7) that what started out as a twelve dimensional integral has been reduced to a much more manageable integration over four variables. The details and validation of this four dimensional integration will be presented in Sec. 3, but first we discuss the change of variables in signal case.

2.3 Changing Variables for Signal *pdf*

To construct the detector level signal *pdf*, which is now a function of the effective couplings $\vec{\mathcal{A}}$, we follow the same procedure as for the background through the second line in Eq.(6) to obtain,

$$P_S(\vec{X}^{\rm R}|\vec{\mathcal{A}}) \;=\; \int P_S(\vec{X}^G|\vec{\mathcal{A}})T(\vec{c}\,|\vec{P}^G)\times\frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|}\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}\prod_{i=1}^{4}dc_i d\vec{p_i}_\perp^G. \tag{13}$$

In contrast to the background however, we now perform the following change of variables,

$$P_S(\vec{X}^{\rm R}|\vec{\mathcal{A}}) = \int P_S(\vec{X}^G|\vec{\mathcal{A}})T(\vec{c}\,|\vec{P}^G)\times\frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|}\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}|\mathbf{J}_S^{\vec{M}}|\cdot d\hat{s}^G dc_1 dM_1^{2G} dM_2^{2G},$$

where again we have implicitly used the delta functions in the transfer function to perform the eight dimensional integration over $\vec{p_i}_\perp^G$. Here $|\mathbf{J}_S^{\vec{M}}|$ is the Jacobian obtained analytically in the change of variables $c_2, c_3, c_4 \to \hat{s}^G, M_1^{2G}, M_2^{2G}$ by using Eq.(10) and the following relation for $\hat{s}^G$,

$$\hat{s}^G = \sum_{i>j} c_i^{-1} c_j^{-1} M_{ij}^{R2}. \tag{14}$$

This allows us to write down the transformation matrix as,

$$\hat{M} = \begin{bmatrix} \dfrac{\partial M_1^{G2}}{\partial c_2} & \dfrac{\partial \hat{s}^G}{\partial c_2} & 0 \\ 0 & \dfrac{\partial \hat{s}^G}{\partial c_3} & \dfrac{\partial M_2^{G2}}{\partial c_3} \\ 0 & \dfrac{\partial \hat{s}^G}{\partial c_4} & \dfrac{\partial M_2^{G2}}{\partial c_4} \end{bmatrix}, \tag{15}$$

from which the Jacobian can be obtained by,

$$|\mathbf{J}_S^{\vec{M}}| = \frac{1}{|det(\hat{M})|}. \tag{16}$$

As discussed in more detail in [1], the 'truth' level $\hat{s}^G$ spectrum for the signal is given by a delta function[1] $\propto \delta(\hat{s}^G - m_h^2)$ (where m_h is the *generated* Higgs mass), enabling us to perform the integration over $d\hat{s}^G$. Thus, we have for the final signal detector level *pdf*,

$$P_S(\vec{X}^{\rm R}|\vec{\mathcal{A}}) = \int P_S(\vec{X}^G|\vec{\mathcal{A}})T(\vec{c}\,|\vec{P}^G)\times\frac{|\mathbf{J}_G^{\vec{P}}|}{|\mathbf{J}_R^{\vec{P}}|}\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}|\mathbf{J}_S^{\vec{M}}|\cdot dc_1 dM_1^{2G} dM_2^{2G}\Big|_{\hat{s}^G=m_h^2}. \tag{17}$$

We note that the delta function in $\hat{s}^G$ introduced additional complications which are computationally non-trivial when including detector resolution effects. This is because the delta function in $\hat{s}^G$ places an additional constraint when performing the M_1^{2G}, M_2^{2G} integration which must be properly taken into account. We discuss these issues in more detail in Sec. 3.7, but first we show how the 12×12 jacobians $\mathbf{J}_G^{\vec{P}}$ and $\mathbf{J}_R^{\vec{P}}$ are computed numerically.

[1]Note that the delta function approximation is taken for the recently discovered $\sim$ 125 GeV Higgs boson, but need not be imposed if a new heavy scalar with a large width is discovered in the future. In this case, the $\hat{s}$ spectrum can be treated similarly to the background case.

2.4 Calculation of the $\mathbf{J}_G^{\vec{P}}$ and $\mathbf{J}_R^{\vec{P}}$ Jacobian Factors

We now turn to the 12×12 Jacobians $\mathbf{J}_G^{\vec{P}}$ and $\mathbf{J}_R^{\vec{P}}$ which take us from the center of mass basis to the 'lepton-smearing basis'. More explicitly these Jacobians define the twelve-dimensional (for known lepton masses) transformation,

$$\vec{X} \equiv (\vec{p}_T, Y, \phi, \hat{s}, M_1, M_2, \vec{\Omega}) \Longrightarrow \vec{P} \equiv (p_{1||}, p_{2||}, p_{3||}, p_{4||}, \vec{p_1}_\perp, \vec{p_2}_\perp, \vec{p_3}_\perp, \vec{p_4}_\perp). \tag{18}$$

Ideally one should simply work out the 12×12 matrix and calculate the discriminant. However, since it involves many boosts and trigonometric functions to perform this non-linear transformation, it is not possible to obtain analytically. Furthermore, since the components of $\vec{P}$ in principle depend on the particular point in $\vec{X}$ this transformation must be obtained for each point in phase space. Therefore we take another approach and calculate the factor numerically.

The Jacobian factors have a simple geometrical interpretation; they can be interpreted as the ratio of infinitesimal volume elements before and after the change of basis. This is illustrated schematically in Fig. 1 where V and V' represent the infinitesimal volume elements around a given point in the two different coordinate systems $\vec{X}$ and $\vec{P}$ respectively. Note that even though the lepton momentum basis is dependent on the particular point in $\vec{X}$, this does not affect the final calculation of the Jacobian since:

- As we scan through the different $\vec{X}$ and transform them to the basis $\vec{P}$, the parallel component directions in $\vec{P}$ line up during the integration. Thus, we can use the same basis for each point in $\vec{X}$ during the integration over $p_{i||}$.
- The freedom of choice in the perpendicular components is irrelevant since the delta function in $\vec{p_i}_\perp$ constrains these directions to be fixed. Without this constraint, one would need to carefully line up the perpendicular directions during the convolution integral.
- The lepton vector basis $\vec{P}$ between different points in the $\vec{X}$ basis are related by simple rotations which leave the volume, and thus the Jacobian, invariant.

These are the key features which make the integration possible and allow us to, point by point in the phase space, numerically build an infinitesimal 12-dimensional cube with volume V in the basis $\vec{X}$ and then transform it to the basis $\vec{P}$ where the volume V' of the resulting hyper-parallelepiped is calculated. Calculating the volume of this 12-dimensional parallelepiped can be done using various readily available algorithms [43] allowing us to obtain the volume V'. We implement a simple algorithm where the hyper-parallelepiped is transformed into a hyper-cube with equal volume V' as follows:

- Choose any vector and pair it with a second vector.
- Subtract out the parallel component of the first vector from the second vector.
- Take a third vector, subtract out the parallel components of the first vector, as well as that of the modified second vector.
- Repeat the process for all remaining sides of the parallelepiped until a hyper-'cube' can be constructed.
- Product of length of all the edges now gives the volume V'.

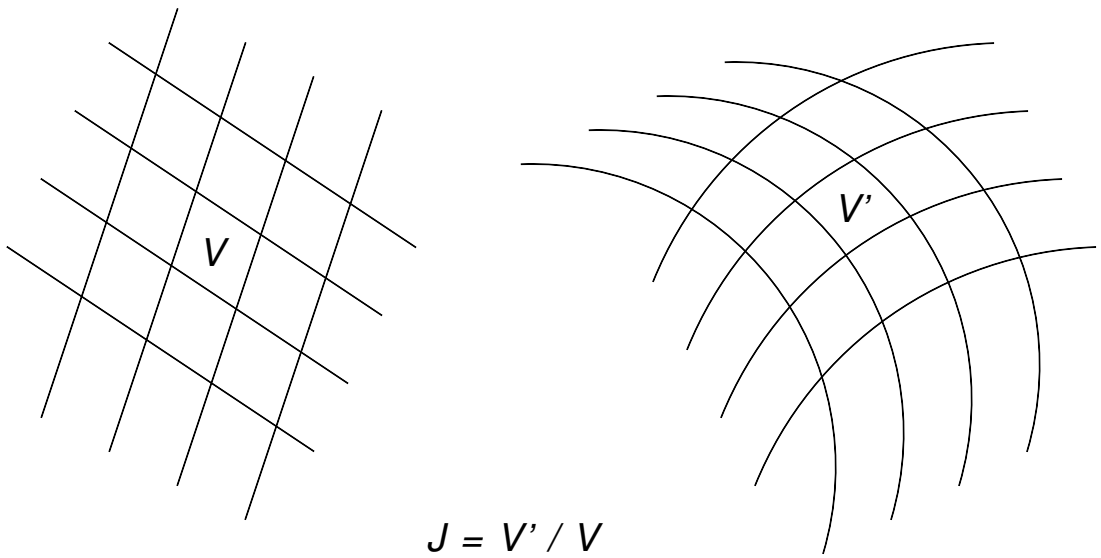

Figure 1: The Jacobian factor can be thought as the ratio of an infinitesimal volume around the given point of interest in the two different bases $\vec{X}$ and $\vec{P}$ labeled by V and V' respectively. Lines on the left in the $\vec{X}$ basis correspond to lines on the right in the $\vec{P}$ basis. The volume V is translated into the volume V' on the right. The Jacobian factor the particular point is therefore $J = V'/V$.

This is equivalent to calculating the determinant of the 12×12 transformation matrix, but conceptually easier to visualize.

This calculation of this Jacobian factor can be validated with toy distributions. For any given test distribution $f(\vec{X})$, we can generate events in two different bases and compare the distribution with one weighted by the Jacobian factor. To see this consider,

$$f(\vec{X})d\vec{X} = f(\vec{X})|\mathbf{J}^{\tilde{\mathbf{P}}}|d\vec{P}, \tag{19}$$

where $\mathbf{J}^{\tilde{\mathbf{P}}}$ represents either $\mathbf{J}_G^{\vec{P}}$ or $\mathbf{J}_R^{\vec{P}}$. Starting from Eq.(19) one can generate events from the left hand side and right hand side separately and then compare the two datasets. They should be identical if the Jacobian has been calculated correctly. Since the Jacobians arise from the change of variable and do not depend on other details of the integration, one has the freedom to choose different toy integrands to validate the calculation of this Jacobian factor. The toy function used is as follows:

$$f(\vec{X}) = \begin{cases} 1, & \text{if } \ 100 \text{ GeV} < \sqrt{\hat{s}} < 140 \text{ GeV}, \\ & 4 \text{ GeV} < M_{1,2} < 100 \text{ GeV}, \quad |\vec{p}_i| < 100 \text{ GeV} \\ & |Y| < 4, \quad |\vec{p}_T| < 100 \text{ GeV} \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

The result of the validation is shown in Fig. 2 where we have decomposed the four lepton system transverse momentum into its components as $\vec{p}_T = (p_T \cos\phi_{4\ell}, p_T \sin\phi_{4\ell})$. We see excellent agreement between events generated in the two different basis for all twelve variables, indicating that the calculation is performed correctly.

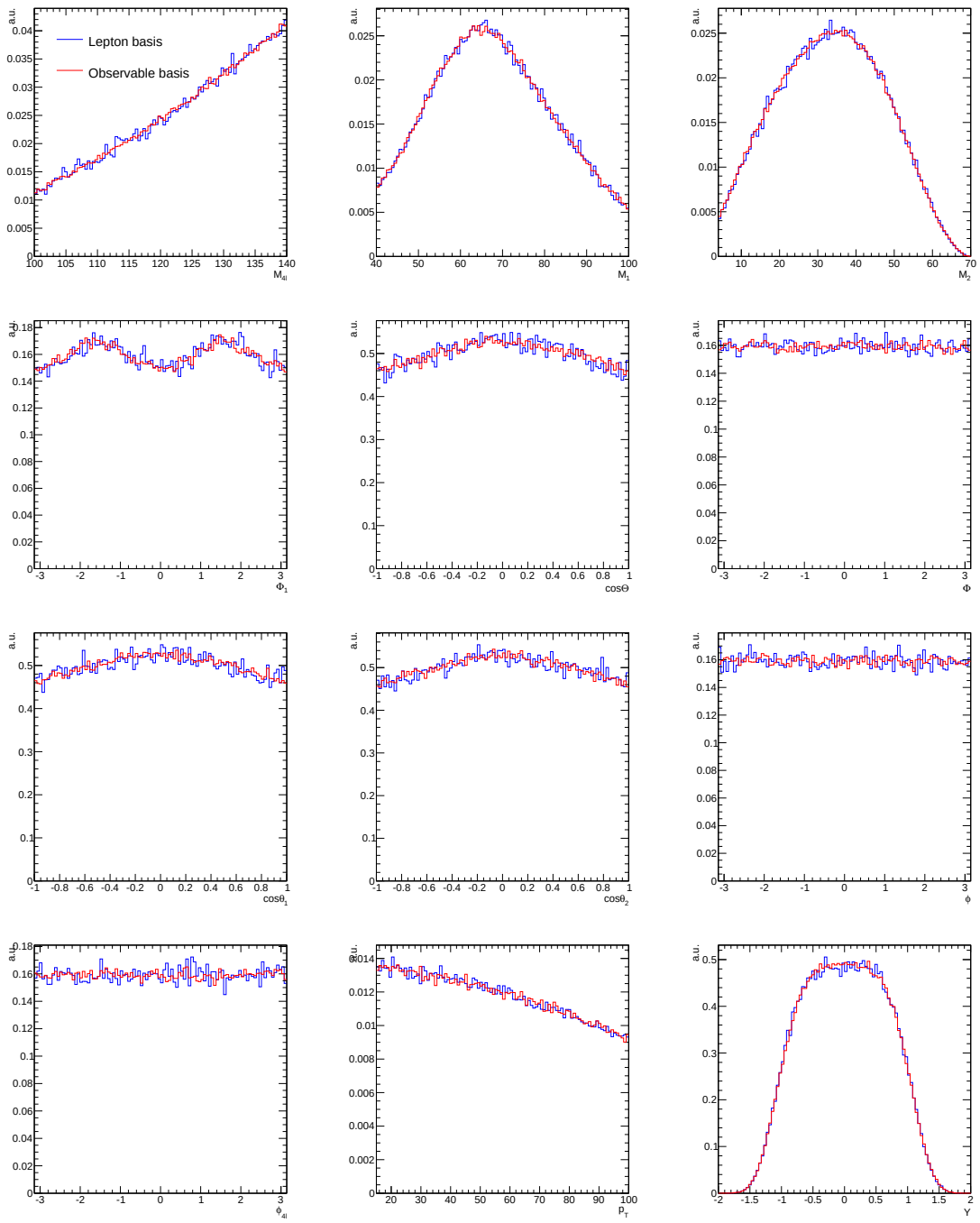

Figure 2: Validation of the 12-dimensional Jacobian using the 'toy' function in Eq.(19). Events are generated uniformly in both bases, and compared to each other by weighting those from the lepton basis $\vec{P}$ by their respective 12×12 Jacobian factor. We see excellent agreement for all variables of interest with any differences due to statistical fluctuations.

3 Integration Over Lepton Momenta

We now turn to how the integration over the lepton momenta is performed in Eq.(7) and Eq.(17). The integral is performed by numerical methods based on Gaussian quadrature [43]. We proceed by first scanning over a grid in the two di-lepton mass directions (M_1^2 and M_2^2)[2] and at each grid point integrating over the remaining c_i smearing factors. Over the grid of mass directions, it is natural that contributions to the final integral are concentrated around certain generator ('truth') level configurations. A strategy is developed to maximize the precision of this procedure during the convolution integration which we now discuss.

3.1 Scan Directions

We observe that in the background case the two mass directions are not correlated. Variation in one mass direction is not dependent on the other mass direction. Therefore a scan in the two mass directions (along M_1 and M_2) is a sufficiently good choice. In the signal case, however, the correlation becomes much stronger due to the narrow width of the resonance and scanning along these directions is no longer optimal. By constraining the mass of the four lepton system, a negative correlation is introduced to the two di-lepton masses. It is thus advantageous to pick a "diagonal direction" as the direction of scanning over the grid. This is done by defining 'mass scan variables' as follow,

$$\begin{aligned} m_+^2 &= M_1^2 + M_2^2 + R_m M_1 M_2, \\ m_-^2 &= (M_1 - M_2)^2, \end{aligned} \tag{21}$$

where R_m is formed using *reconstructed* level di-lepton masses as,

$$R_m \equiv \frac{{m_{13}^R}^2 + {m_{14}^R}^2 + {m_{23}^R}^2 + {m_{24}^R}^2}{m_{12}^R m_{34}^R}, \tag{22}$$

and m_{ij}^R are the invariant masses formed by reconstructed lepton i and lepton j. Reconstructed masses are fixed and taken as input in the integration process. By doing the scan in these directions, we obtain an additional Jacobian factor which is easily calculated to be,

$$J_m = (R_m + 2)(M_1^2 + M_2^2). \tag{23}$$

An example of grid lines along this modified direction is shown in Fig. 3.

3.2 Non-Uniform Grid Spacing

In addition to using the mass scan directions discussed above, we also use a non-uniform grid for further optimization. Due to the nature of the contributions concentrated around a certain 'truth level' parameter point, we choose the grid to be more dense in the center. For a uniform grid, one can write the location $\bar{x}_i$ of each grid point as,

$$\begin{aligned} \bar{x}_i &= x_o + \delta x_i, \\ \delta x_i &\equiv \frac{\Delta x}{N_{\text{grid}}} \left(i - \frac{1}{2} N_{\text{grid}} \right), \end{aligned} \tag{24}$$

[2] In what follows we will no longer explicitly write the superscript G for generator level observables which serve as our variables of integration. We will however continue explicitly write the superscript R for reconstructed level variables which are fixed quantities in the integration procedure.

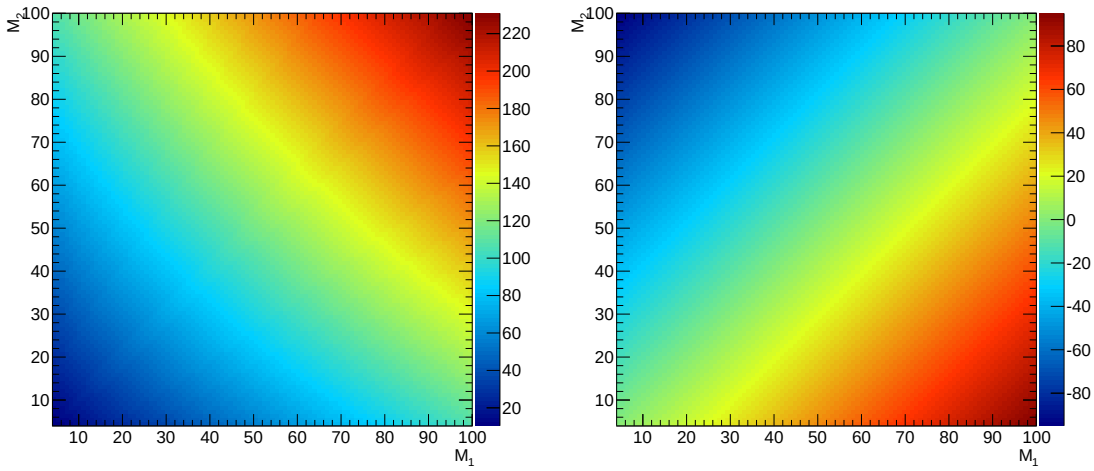

Figure 3: Contour of m_+ (left) and m_- (right) for one example event as a function of M_1 and M_2. The contour for m_+ changes from event to event, but is always roughly diagonal.

where x_o is the center point of the scan within the grid (which we attempt to place near the 'true' point), Δx is the distance between the leftmost point and rightmost point of the grid, and N_{grid} is the number of grid points.

When allowing for non-uniform grid spacing, Eq.(24) for the location of the grid point must be modified. The grid spacing can be characterized by an "attractor" parameter A_S. With this attractor we now define the new modified grid point locations $\tilde{x}_i$ as,

$$\tilde{x}_i = x_o + \frac{\delta x_i(|A_S \delta x_i| + 1)}{|A_S \Delta x|/2 + 1}. \tag{25}$$

The linear spacing is now modified to be quadratic, with center point and end points the same as before. Larger $|A_S|$ values result in denser grid spacing near the center with linear spacing recovered when $A_S = 0$ giving $\tilde{x} \to \bar{x}$. This behavior is illustrated in Fig. 4. How the central point x_o is chosen will be discussed in Sec. 3.4.

3.3 Modified 2nd Order Newton-Cotes Formula

With the modification of grid spacing, it is necessary to derive the equivalent of Newton-Cotes formula [43] for a non-uniform grid. We work with 2nd order closed integration where to each interval we assign three points to which we fit a second-order polynomial and then obtain the integral (area). Furthermore, the area can be written as a weighted sum of the points used in the fit. This can be seen by considering three points located at $-\delta, 0$, and $+\delta$ with height $f(-\delta), f(0)$ and $f(+\delta)$ respectively. In the case of uniform gird spacing the area (of the function $f(x)$) estimator can then be written as the following [43],

$$I(-\delta, +\delta) = 2\delta \, \frac{f(-\delta) + 4f(0) + f(\delta)}{6}. \tag{26}$$

For non-uniform grid spacings, Eq.(26) needs to be re-derived. Since the integral does not depend on the absolute value of the x-axis coordinates, we can take the liberty to pick the center point at zero, and the other two located at $x = -\delta_-$ and $x = \delta_+$ for the left and right endpoints of the

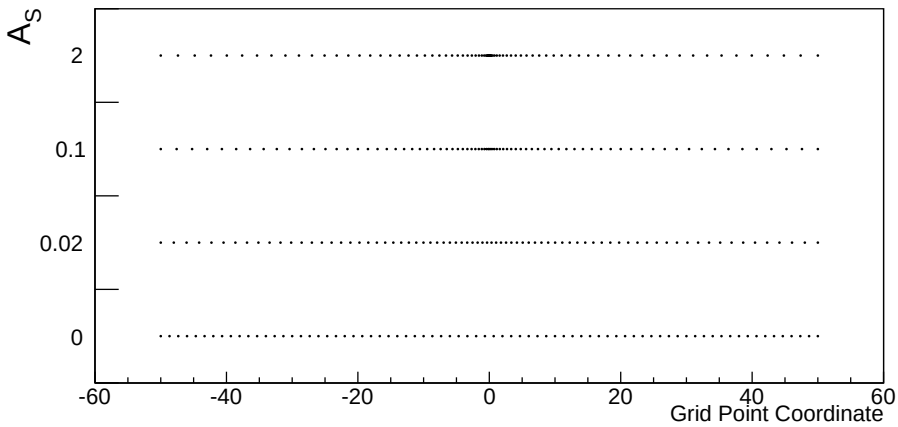

Figure 4: Demonstration of the mass grid attractor. Different sets of grid points are plotted in each line with varying mass grid attractor strength. With attractor strength set to zero, we recover uniform spacing (bottom), while large values of attractor strength causes points to be concentrated near the center (top).

interval respectively. Integrands at the different points can be written as $f(-\delta_-), f(0), f(\delta_+)$. For non-uniform spacing the integral estimator is modified to,

$$I(-\delta_-, \delta_+) = \frac{1}{3}A(\delta_+^3 + \delta_-^3) + \frac{1}{2}B(\delta_+^2 - \delta_-^2) + C(\delta_+ + \delta_-), \tag{27}$$

where A, B and C are the coefficients of the 2'nd order polynomial which is used in the fit to the points in a given interval and defined as,

$$y(x) = Ax^2 + Bx + C. \tag{28}$$

In terms of the three points in the interval the coefficients are obtained via the matrix equation,

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \frac{1}{\delta_+\delta_-(\delta_+ + \delta_-)} \begin{bmatrix} \delta_- & -\delta_+ - \delta_- & \delta_+ \\ \delta_-^2 & \delta_+^2 - \delta_-^2 & -\delta_+^2 \\ 0 & \delta_+\delta_-(\delta_+ + \delta_-) & 0 \end{bmatrix} \begin{bmatrix} f(\delta_+) \\ f(0) \\ f(\delta_-) \end{bmatrix}. \tag{29}$$

Together Eq.(27) and Eq.(29) define the final reweighing coefficients to use for the non-uniform gird integration. When $\delta_- = \delta_+$ they reduce to the linear formula in Eq.(26). The analogous two-dimensional formula can be trivially obtained by using Eq.(27) and Eq.(29) multiplicatively on the two dimensions.

3.4 Central Grid Point Optimization

During the mass scan, we do not know *a priori* where the location of the maximum contribution to the final integral will be. If we do not center our mass grid scan close to the maximum contribution point, there is a chance part of the contributions will be missed. It is especially crucial when we consider non-uniform grid spacings to increase efficiency in mass integration.

We therefore employ a simple numerical algorithm to find where the maximum contribution point is. It is outlined as follows:

- For the background we start from a best-guess value of reconstructed masses M_1^R and M_2^R while choosing a reasonable window size.
- For signal we choose the best-guess value to be $(\sqrt{\hat{s}/\hat{s}^R})M_1^R$ and $(\sqrt{\hat{s}/\hat{s}^R})M_2^R$ again choosing a reasonable window size.
- We put a coarse grid in this window centered at the current best guess value and evaluate the integrand (integrals of c_1 and c_3) at each point.
- The integration is carried out repeatedly with the central point updated each time to the point which gives the largest contribution.
- We then reduce the window size by a fraction and repeat the process.
- The process terminates when the point with the largest contribution lies within the 5% most central grid points after which we adjust the grid window.
- The final best guess value is used as the center point for constructing the mass grid.

During the integration we also keep track of the RMS of the integral contribution in units of number of grid points. If the RMS is found to be less than four grid points in either direction, the integration is repeated with a reduced window. This ensures us of a grid center that is sufficiently close to the 'truth level' point with maximum contribution.

A demonstration of the whole process on one example signal event is shown in Fig. 5. Each bin in the plot is one grid point which has been scanned over and the color indicates the inner integral from that given bin. On the left we have the usual scanning directions (along M_1 and M_2) and on the right we show the scan along the modified diagonal directions (along m_+ and m_-) as explained in Sec. 3.1. In the top two panels the scanning is done without implementing the mass grid attractor and without the central grid point optimization. In the middle two panels the mass grid attractor is turned on, but without central grid point optimization. In the bottom panels both the mass grid attractor and central grid point optimizations are turned on. With all optimizations 'turned on' the amount of grid points with significant contributions is greatly enhanced with decreasing correlation between the two scanning directions. This provides us increasing precision and stability during the integration.

3.5 Differential Cross Section Expansion for Background

We now turn to the integration over the smearing factors (c_i) in Eq.(7) and Eq.(17). As will be discussed in more detail in the next section, we wish to further control precision by applying an adaptive integration method for the integrals over these variables. In the background case, at each grid point for the masses, there is a two-dimensional integral over c_1 and c_3 which must be performed. Since a brute force recursive integration method in two dimensions would take too much computing resources, we start by making a few observations.

We first note that in the final integration expression for the background (Eq.(7)), the transfer function can be factorized into terms for the first lepton pair and terms for second lepton pair. This assumes the transfer function is a product of functions for each lepton individually which is a very good approximation for the CMS and ATLAS detectors [41, 42] once standard lepton selection criteria are imposed[3]. Furthermore, the range of smearing factors is considered to be

[3]If the approximation of sufficiently separated leptons does not hold, the integration procedure can not be carried out as formulated and must be altered. We do not explore these complications associated with leptons with very small separations and can avoid them by simply imposing a minimum $m_{\ell\ell}$ cut between any two leptons.

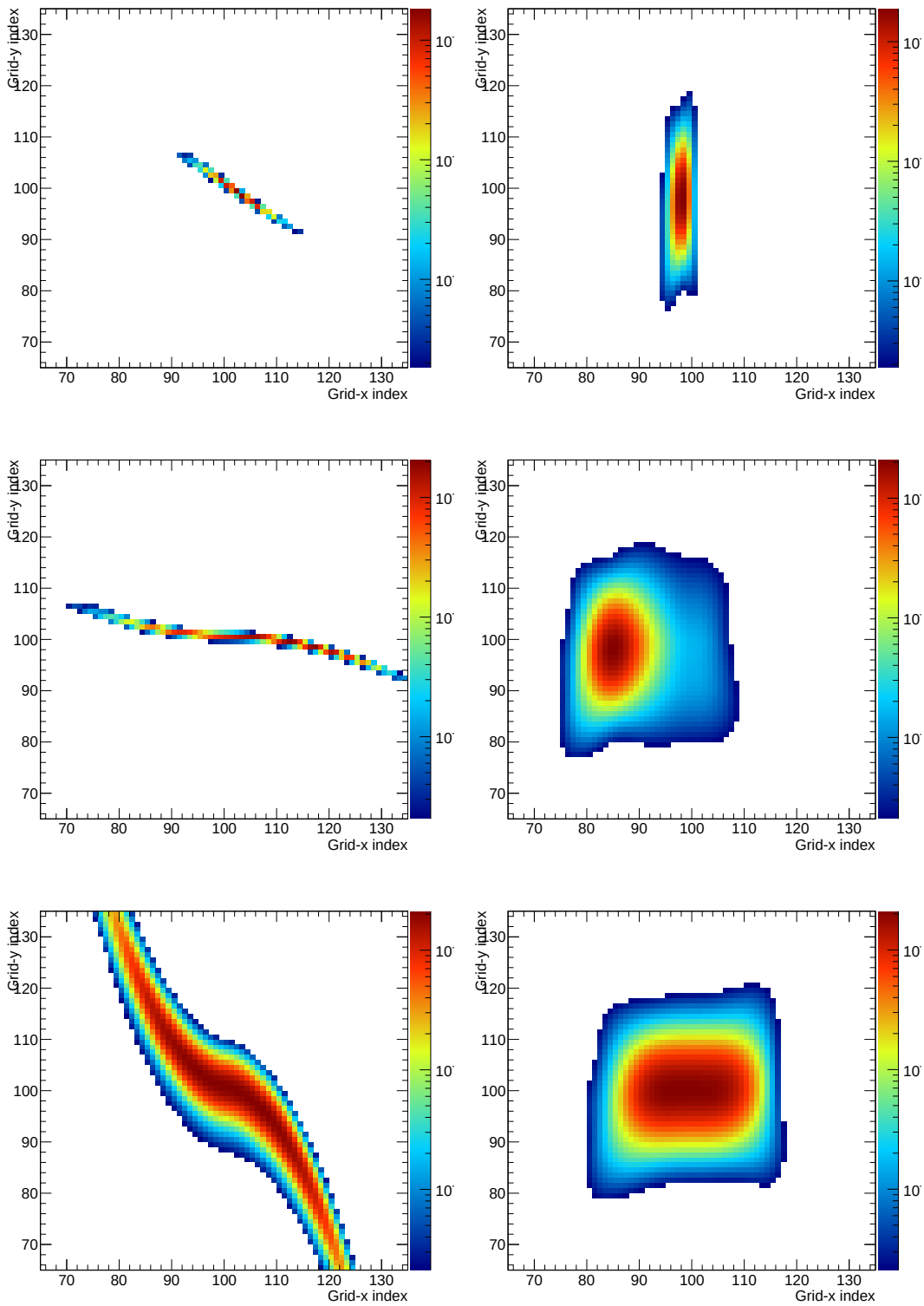

Figure 5: Demonstration of mass grid attractor and scan directions during the integration. Each bin is one grid point, where the color indicates the value of the inner integral. In the top two plots the attractor strength is set to zero. In the middle plots we have chosen a moderate attractor strength while in the bottom plots the center grid point optimization is turned on. Plots on the left are along the scanning directions M_1, M_2 while on the right we show the modified directions m_+, m_-.

narrow since the probability that the lepton momentum is smeared far from its 'true' value is very small [40–42]. This allows us to write the rest of the integrand, apart from the transfer function, as a series expansion in smearing parameters (about the 'true' value of one) as follows,

$$\begin{aligned}
\mathcal{F}_B(\vec{X}^R) &\equiv P_B(\vec{X}^G)|\mathbf{J}_G^{\vec{P}}|\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}|\mathbf{J}_B^{\vec{M}}| \\
&\approx \mathcal{F}_B\Big|_{c_1=c_3=1} + \frac{\partial \mathcal{F}_B}{\partial c_1}\Big|_{c_1=c_3=1}(c_1-1) + ... \\
&\equiv B_0 + B_1(c_1-1) + B_3(c_3-1) + B_{11}(c_1-1)^2 \\
&+ B_{13}(c_1-1)(c_3-1) + B_{33}(c_3-1)^2 + ..., \qquad (30)
\end{aligned}$$

where $c_i = 1$ corresponds to the 'truth level' value for the smearing factors and the differential cross sections are implicitly absorbed into the B_{ij} coefficients. We can then factorize the inner integral into a sum of products of single integrals. Thus we have,

$$\begin{aligned}
P_B(\vec{X}^R) &= \frac{1}{|\mathbf{J}_R^{\vec{P}}|}\int P_B(\vec{X}^G)T(\vec{c}|\vec{P}^G)|\mathbf{J}_G^{\vec{P}}|\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}|\mathbf{J}_B^{\vec{M}}|dc_1dc_3dM_1^2dM_2^2 \\
&= \frac{1}{|\mathbf{J}_R^{\vec{P}}|}\int \left(P_B(\vec{X}^G)T(\vec{c}|\vec{P}^G)|\mathbf{J}_G^{\vec{P}}|\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}|\mathbf{J}_B^{\vec{M}}|dc_1dc_3\right)dM_1^2dM_2^2,
\end{aligned}$$

where the inner integral can now be expanded as,

$$\begin{aligned}
&\int P_B(\vec{X}^G)T(\vec{c}|\vec{P}^G)|\mathbf{J}_G^{\vec{P}}|\frac{|\mathbf{J}_G^{\vec{c}}|}{|\mathbf{J}_R^{\vec{c}}|}|\mathbf{J}_B^{\vec{M}}|dc_1dc_3 \\
\approx\ & B_0\int T_{12}(\vec{c}|\vec{P}^G)(c_1-1)^0dc_1\int T_{34}(\vec{c}|\vec{P}^G)(c_3-1)^0dc_3 \\
+\ & B_1\int T_{12}(\vec{c}|\vec{P}^G)(c_1-1)^1dc_1\int T_{34}(\vec{c}|\vec{P}^G)(c_3-1)^0dc_3 \\
+\ & B_3\int T_{12}(\vec{c}|\vec{P}^G)(c_1-1)^0dc_1\int T_{34}(\vec{c}|\vec{P}^G)(c_3-1)^1dc_3 \\
+\ & B_{11}\int T_{12}(\vec{c}|\vec{P}^G)(c_1-1)^2dc_1\int T_{34}(\vec{c}|\vec{P}^G)(c_3-1)^0dc_3 \\
+\ & B_{13}\int T_{12}(\vec{c}|\vec{P}^G)(c_1-1)^1dc_1\int T_{34}(\vec{c}|\vec{P}^G)(c_3-1)^1dc_3 \\
+\ & B_{33}\int T_{12}(\vec{c}|\vec{P}^G)(c_1-1)^0dc_1\int T_{34}(\vec{c}|\vec{P}^G)(c_3-1)^2dc_3 + ... \\
\equiv\ & B_0F_{12}^{(0)}F_{34}^{(0)} + B_1F_{12}^{(1)}F_{34}^{(0)} + B_3F_{12}^{(0)}F_{34}^{(1)} \\
+\ & B_{11}F_{12}^{(2)}F_{34}^{(0)} + B_{13}F_{12}^{(1)}F_{34}^{(1)} + B_{33}F_{12}^{(0)}F_{34}^{(2)} + ... , \qquad (31)
\end{aligned}$$

and we have defined the integrals,

$$F_{ij}^{(n)} \equiv \int T_{ij}(\vec{c}|\vec{P}^G)(c_i-1)^n dc_i. \qquad (32)$$

The transfer functions for the first and second lepton pairs are denoted as $T_{12}(\vec{c}|\vec{P}^G)$ and $T_{34}(\vec{c}|\vec{P}^G)$ respectively. Thus we see that the two-dimensional integral over the smearing parameters has been reduced to a product of single integrals which are computationally easier to control. With this

procedure one can increase the expansion order if higher precision is required. The B_{ij} coefficients in front of each $F_{ij}^{(n)}$ term can be obtained by finding an approximate two-dimensional polynomial to the target function over the whole integration range. In the current implementation it is done by picking a few points as representative and finding a polynomial that goes through all the points.

3.6 Recursive Integration

The integration over smearing factors c_1 and c_3 is done by a recursive algorithm [43] which provides a handle on precision. The algorithm begins by first splitting the integration range into multiple segments and then applying the recursive algorithm to each segment with some pre-specified tolerance level ε. More explicitly the recursive algorithm proceeds as follows:

- Apply second order Newton-Cotes quadrature to the whole range to obtain a first order approximation for the integral which we label I_0.
- We then split the segment in half and apply the approximation on each of the half-segments to obtain a second estimate $I_1 + I_2$.
- Next we estimate integration error δI by comparing I_0 and $I_1 + I_2$.
- If $|\delta I| < \varepsilon$, we terminate the algorithm and use $I_1 + I_2$ as the integral.
- Otherwise, we repeat the procedure on each of the two half-segments and require that each satisfy a tolerance level of $\varepsilon/2$.

This procedure ensures that the overall error is at most ε, assuming that the error estimation is reasonable, which when the function does not vary too rapidly is a good approximation. This is aided by the fact that the integral is split into smaller intervals where the function varies less and then a sum is taken over all the intervals.

To assess whether the estimate of the error on the integration is reasonable, we first suppose the step size is small enough that there is not a drastic change of landscape inside each integration segment. Since the estimator we use is a fourth order method, the difference between the actual integral and the estimator can be written as [43],

$$\int_{segment} f(x)dx = I_0 + kh^4 + O(h^5), \tag{33}$$

where h is the size of segment and k is a constant characteristic of the method and independent of h. Applying the same formula to the total of the two half-segments, we arrive at the following,

$$\int_{segment} f(x)dx = I_1 + I_2 + 2k\left(\frac{h}{2}\right)^4 + O(h^5). \tag{34}$$

Assuming that the term inside $O(h^5)$ is negligible, by comparing the two formulae we obtain an estimate of integration error to be,

$$\delta I \equiv 2k\left(\frac{h}{2}\right)^4 = \frac{1}{7}\left(I_1 + I_2 - I_0\right). \tag{35}$$

In Fig. 6 we show an example of the integration using the adaptive method on two test functions. In the left panel we perform integration on a second-order polynomial $f(x) = \frac{1}{75}x^2$, where the method is exact. In this case the grid spacings are always the same. In the right panel the integrand is modified to be $f(x) = \frac{1}{75}x^2 + e^{-\frac{1}{2}x^2}$. We can see that in places where functions are rapidly varying, a greater number of points are used.

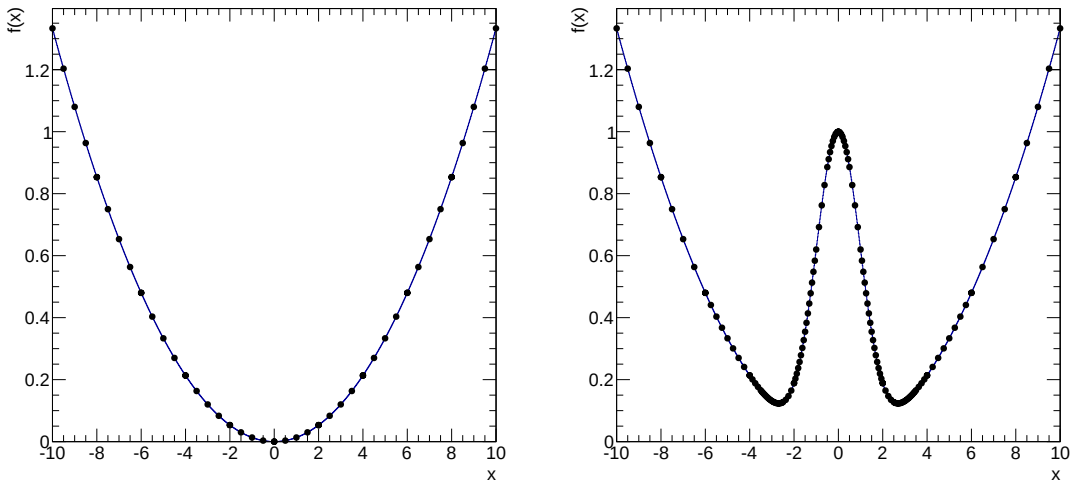

Figure 6: Illustration of recursive integration. On each plot, the line shows true function, and dots are points evaluated for the integration. In the left panel we perform integration on a second-order polynomial $f(x) = \frac{1}{75}x^2$, where the method is exact and therefore, the grid spacings are always the same. In the right panel the integrand is modified to be $f(x) = \frac{1}{75}x^2 + e^{-\frac{1}{2}x^2}$. We can see that in places where the function is rapidly varying, more points are used.

3.7 Signal Case Complications

As discussed in previous sections, in the signal case an anti-correlation between the two lepton pairs is induced due to the presence of the narrow-width resonance. We can see the effect on the di-lepton mass due to the narrow-width approximation in Fig. 7. In both plots the inner integral is plotted for the exact same event as a function of M_1 and M_2. On the left we do not enforce narrow-width approximation while on the right we require that $\sqrt{\hat{s}} = 125$ GeV. We see the shape is strongly elongated for the case with narrow-width approximation.

By using a delta function to model width of resonance, there is one less dimension to integrate over as compared to the background case. While this makes it easier computationally in one respect, an additional complication arises since we have to integrate along a trajectory in which $\hat{s}$ is kept constant. Specifically, for a given value of c_1 we have to calculate the corresponding value of the other smearing parameter such that M_1, M_2, and $\hat{s}$ are held constant. It is easy to keep the di-lepton masses constant since the integral over c_1 is the innermost integral and,

$$c_1 c_2 M_1^2 = {m_{12}^R}^2, \tag{36}$$

which allows us to simply choose $c_2 = {m_{12}^R}^2/M_1^2 c_1$ in order to satisfy the necessary conditions (and similarly for the second lepton pair). Satisfying the last condition where $\hat{s}$ is kept constant involves a less trivial calculation. We begin from the equation,

$$\sum_{i>j} c_i^{-1} c_j^{-1} m_{ij}^{R\,2} = \hat{s}^2, \tag{37}$$

from which we obtain a quadratic equation for c_3 given by,

$$\left(R_{34}{m_{14}^R}^2 + c_1^2 R_{34} R_{12} {m_{24}^R}^2\right) c_3^2 - (m_h^2 - M_1^2 - M_2^2) c_1 c_3 - \left({m_{13}^R}^2 + c_1^2 R_{12} {m_{23}^R}^2\right) = 0, \tag{38}$$

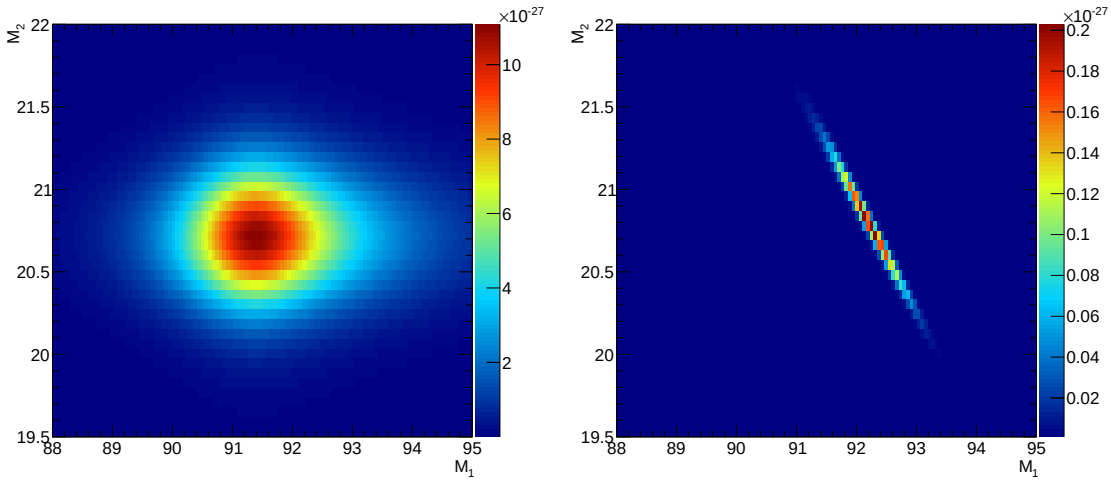

Figure 7: Demonstration of the effect on di-lepton masses due to narrow-width approximation. In both plots the inner integral as a function of M_1 and M_2 is plotted for the exact same event. On the left we do not enforce narrow-width approximation while on the right we require that $\sqrt{\hat{s}} =$ 125 GeV. We see shape is strongly elongated for the case with narrow-width approximation.

where $R_{12} = M_1^2/{m_{12}^R}^2$ and $R_{34} = M_2^2/{m_{34}^R}^2$. One condition where a solution exists for c_3 given c_1 value is as follows,

$$\left[\left(m_h^2 - M_1^2 - M_2^2\right)^2 - 4R_{34}R_{12}{m_{14}^R}^2{m_{23}^R}^2 - 4R_{12}R_{34}{m_{13}^R}^2{m_{23}^R}^2\right]c_1^2$$
$$-4R_{34}{m_{14}^R}^2{m_{13}^R}^2 - 4R_{12}^2R_{34}{m_{24}^R}^2{m_{23}^R}^2c_1^4 \geq 0, \tag{39}$$

which is a quadratic function of c_1^2. We first make a few observations based of Eq.(39). First of all, the quadratic coefficient is negative definite, indicating that the parabola is curving downwards. The constant term is also negative definite. If there exists a positive solution for c_1^2, both solutions are positive. We also require that the linear coefficient is positive which ensures that if there is a solution, there exists a positive solution. To summarize, the two conditions that need to be met in order for solutions of the smearing parameters to exist that will meet all mass requirements are given by,

$$\left(m_h^2 - M_1^2 - M_2^2\right)^2 - 4R_{34}R_{12}{m_{14}^R}^2{m_{23}^R}^2 - 4R_{12}R_{34}{m_{13}^R}^2{m_{23}^R}^2$$
$$> 8R_{12}R_{34}m_{13}^Rm_{14}^Rm_{23}^Rm_{24}^R \quad (> 0). \tag{40}$$

The first equality is the necessary condition for Eq.(39) to be true combined with the condition that the linear coefficient is to be positive. If these conditions are met, all allowed solutions form an ellipsoidal contour in the positive quadrant of the $c_1 - c_3$ plane. Some example ellipses are shown in Fig. 8.

Due to the shape, we need to integrate different regions of the ellipse separately. To do this we "slice" this ellipse into four regions. First, two points are picked by calculating the average value of c_1 between the leftmost point of the ellipse and the smaller of two points where $dc_3/dc_1 =$ 0. Similarly we can also define another point using the rightmost point and the larger of two points where $dc_3/dc_1 = 0$. The four regions are thus defined by the two horizontal segments between the two aforementioned points and two vertical segments on the two sides. An example is

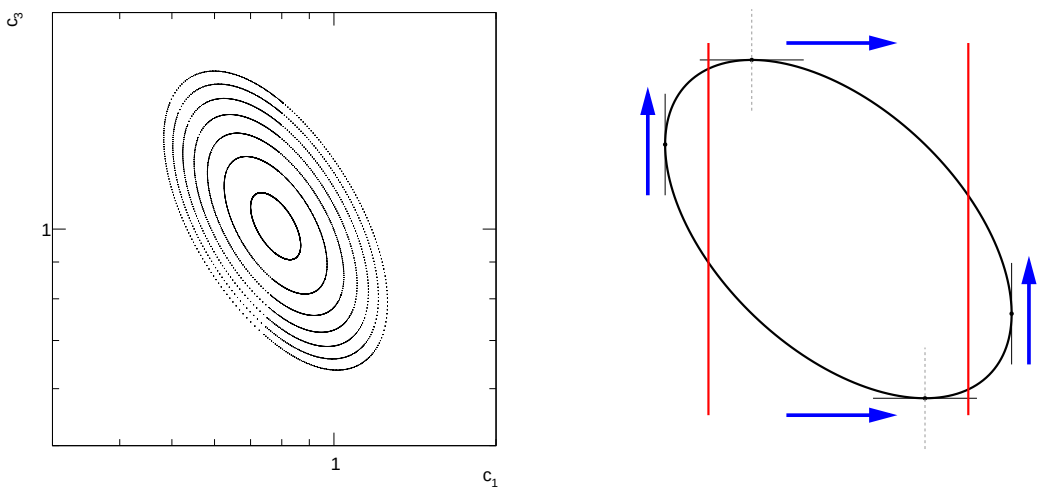

Figure 8: Examples of $c_1 - c_3$ ellipses for a few mass grid points on an example event are shown in the left panel where each contour represents one mass grid point. Dots represent possible solutions to all mass constraints. For each grid point all possible solutions form an ellipsoidal shape, as expected. In the right we show the slicing of ellipse while doing the integral. The four extreme points (left, right, up, down) are calculated and the curve is divided into four regions by the red lines going vertically between the extreme points. In order to avoid artificial infinities during integration we integrate along the horizontal direction for the two segments in the middle and vertically upwards along c_3 for the two segments on the side shown by blue arrows.

shown in Fig. 8. We integrate the horizontal segments normally over c_1. For the vertical segments we flip the role of c_1 and c_3 and integrate along the c_3 direction, adjusting c_1 so that all mass requirements are met. This is necessary as the signal Jacobian factor $|\mathbf{J}_S^{\vec{M}}|$ diverges at the edge of ellipse where the direction of curve approaches vertical.

The complete validation of the full convolution procedure can be found in [1, 39, 40].

4 Constructing Likelihoods and Maximization

To obtain the final likelihoods we must first normalize the detector level differential cross sections to obtain a proper *pdf*. Once a *pdf* is obtained we can go on to construct the final likelihoods with which multi-dimensional parameter extraction can be perfumed. In this section we describe how the normalization is performed and sketch how parameter extraction can be done.

4.1 Normalization

As described in the previous section, it is necessary to normalize the *pdf* properly in order to construct the likelihoods. The normalization can be calculated by integrating the $P(\vec{X}^R|\vec{\mathcal{A}})$ over reconstructed level configurations $\vec{X}^R$:

$$N(\vec{\mathcal{A}}) \equiv \int P(\vec{X}^R|\vec{\mathcal{A}}) d\vec{X}^R. \tag{41}$$

The evaluation of the convolution integral $P(\vec{X}^R|\vec{\mathcal{A}})$ is relatively computationally intensive and it is not possible to repeat it many different times to get a numerical average of the overall

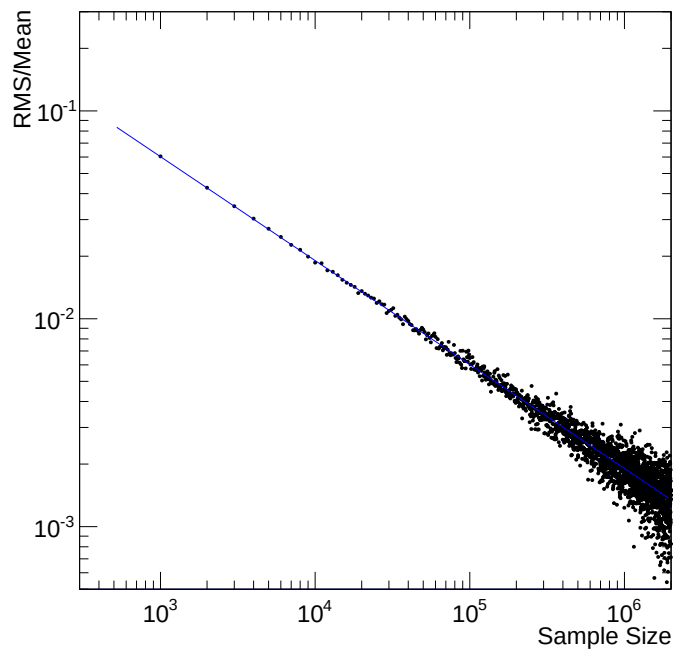

Figure 9: Controlling normalization precision

normalization via Monte-Carlo methods. It is also non-trivial to calculate it via numerical methods as there are 12 dimensions and there is no simple way to cut down the dimensions. We can, however, evaluate the normalization differently and avoid the numerically intensive part as follows,

$$\begin{aligned} N(\vec{\mathcal{A}}) &= \int P(\vec{X}^R|\vec{\mathcal{A}}) d\vec{X}^R \\ &= \int \left(\int P(\vec{X}^G|\vec{\mathcal{A}}) T(\vec{X}^R|\vec{X}^G) d\vec{X}^G \right) d\vec{X}^R \\ &= \int P(\vec{X}^G|\vec{\mathcal{A}}) \left(\int T(\vec{X}^R|\vec{X}^G) d\vec{X}^R \right) d\vec{X}^G \\ &\equiv \int P(\vec{X}^G|\vec{\mathcal{A}}) \bar{\epsilon}(\vec{X}^G) d\vec{X}^G, \end{aligned} \tag{42}$$

where in the last step we define the "average efficiency" given a certain generator level configuration as $\bar{\epsilon}(\vec{X}^G)$. It is the average efficiency a certain generator-level event will survive all analysis cuts. This way we avoid the computationally difficult parts and a straightforward Monte-Carlo algorithm is sufficiently precise. In Fig. 9 we show an example of convergence of the normalization calculation as a function of sample count and CPU time. We can reach a precision of 0.1% in a few CPU-hours of run time for all components.

4.2 Signal Likelihood for Arbitrary Model Point

Due to the simple (polynomial) dependance of the signal differential cross section on the lagrangian parameters, it is possible to pre-calculate different pieces of the detector level likelihood separately, and later combine them in a trivial way to get the final detector-level likelihood. The polynomial

dependence (quadratic in the simplest case [1]) allows us to write the generator-level differential cross section in the form

$$P_S(\vec{X}^G|\vec{\mathcal{A}}) = \sum_i f_i(\vec{\mathcal{A}}) P_S^i(\vec{X}^G), \tag{43}$$

where the index i runs over different terms in the expression (see [1] for description of the different terms) and $f_i(\vec{\mathcal{A}})$ is some polynomial function of the lagrangian parameters. The convolution integral can then be done on pieces $P_S^i(\vec{X}^G)$ which are independent of $\vec{\mathcal{A}}$ and later combined together:

$$\begin{aligned} P(\vec{X}^R|\vec{\mathcal{A}}) &= \int P(\vec{X}^G|\vec{\mathcal{A}}) T(\vec{X}^R|\vec{X}^G) d\vec{X}^G \\ &= \int \sum_i f_i(\vec{\mathcal{A}}) P_S^i(\vec{X}^G) T(\vec{X}^R|\vec{X}^G) d\vec{X}^G \\ &= \sum_i \left(\int P_S^i(\vec{X}^G) T(\vec{X}^R|\vec{X}^G) d\vec{X}^G \right) f_i(\vec{\mathcal{A}}). \end{aligned} \tag{44}$$

In the case of current analysis where we have calculated up to leading order in these couplings, the differential cross section is a second-order polynomial. The same structure applies to the *pdf* normalization, which is also a polynomial of parameters we want to measure:

$$\begin{aligned} N(\vec{\mathcal{A}}) &= \int P(\vec{X}^G|\vec{\mathcal{A}}) \bar{\epsilon}(\vec{X}^G) d\vec{X}^G \\ &= \sum_i \left(\int P_S^i(\vec{X}^G) \bar{\epsilon}(\vec{X}^G) d\vec{X}^G \right) f_i(\vec{\mathcal{A}}). \end{aligned} \tag{45}$$

Once we have the coefficients pre-calculated, it's straightforward and fast to combine into final likelihood for arbitrary values of parameters of interest. We can obtain the background likelihood in a similar manner except the overall normalization does not depend on any undetermined parameters which simples the procedure. The final signal plus background can then be constructed as described in [1, 39, 40].

4.3 Maximization of Likelihood and Parameter Extraction

There are various ways one can go on to perform parameter extraction once the likelihood is constructed. One can for example simply scan the likelihood for each parameter point as done in [39]. We instead implement a maximization procedure based on the MINUIT [44] function minimization code which is incorporated into our framework in order to find the maximum of the likelihood. More specifically, once the likelihood $L(\vec{\mathcal{A}})$ for a particular dataset is obtained, the maximization procedure is utilized in order to obtain the value of the parameters which maximizes the likelihood, which we label $\hat{\mathcal{A}}$. Thus $\hat{\mathcal{A}}$ represents the most likely value of $\vec{\mathcal{A}}$ for a given dataset which is schematically represented as,

$$\frac{\partial L(\vec{\mathcal{A}})}{\partial \vec{\mathcal{A}}}\Big|_{\vec{\mathcal{A}}=\hat{\mathcal{A}}} = 0. \tag{46}$$

One important feature of the procedure is that the computationally intensive component of evaluating the likelihood only needs to be done for the events in the final dataset used in the fit for a given experiment. Therefore the computationally expensive pieces can be calculated on the

Finding the Global Maximum

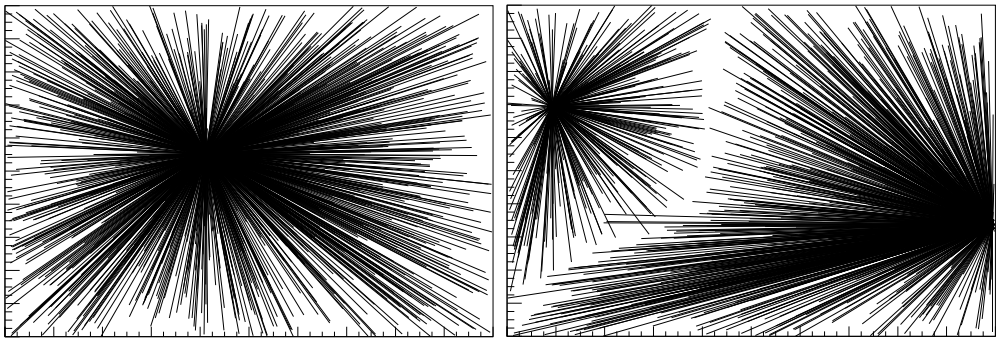

Figure 10: 'Arrow plots' showing the convergence to the point which maximizes the likelihood starting from a random initial point. In these the tail of the arrow is at the initial point while the head is at the end point to which the fit converges. On the left we see that only one end point is found for all initial values. On the right we see the appearance of two endpoints which depend on the initial value indicating the appearance of multiple maxima.

computing grid prior to the analysis of the data, and the fits for the parameter extraction itself is then completed within a few seconds. This allows for a great deal of flexibility when fitting the undetermined parameters.

In practice the maximization in Eq.(46) is done by starting from some random initial point in the parameter space and utilizing the built in algorithm in the MINUIT minimization code to efficiently find the maximum. Of particular importance in this step is ensuring that the point in parameter space that this procedure converges to is actually the *global maximum* and not simply a local maximum, as the statistical fluctuations of a particular dataset can lead to the appearance of multiple local maxima in the likelihood. This can lead to biases or imprecise estimations of the undetermined parameters.

We illustrate this effect in Fig. 10 where we show 'arrow plots' for an example two-dimensional fit to two different datasets containing the same number of events and same 'true' value for the undetermined parameters. We show a large number of arrows whose tails begin at some initial point in a two dimensional parameter space and whose heads point to the final point reached in the maximization scan. On the left we see the same endpoint is reached regardless of the initial starting point indicating there is a clear global maximum. On the right we see two separate accumulations to which the arrow heads point indicating two local maxima. We have carefully accounted for this effect in our maximization procedure and find a very high convergence rate in general ($\gtrsim 99\%$) to the global maximum of the likelihood. Various demonstrations of the parameter extraction can be found in [34, 36] at generator level and inin [39, 40] for detector level.

To quantify the uncertainty on the extracted value $\hat{\mathcal{A}}$ we perform a large number of pseudo-experiments $\mathcal{N}$ each containing N events and perform the maximization for each pseudo-experiment. A distribution for $\hat{\mathcal{A}}$ is obtained with a spread σ and average value $\bar{\mathcal{A}}$. The true value $\mathcal{A}_o$ will sit within some interval of the extracted value $\hat{\mathcal{A}}$ for a given pseudo experiment and as the number of pseudo experiments is taken to infinity the average value of $\hat{\mathcal{A}}$ will converge to the true value; i.e. $\bar{\mathcal{A}} \to \mathcal{A}_o$ as $\mathcal{N} \to \infty$. Further discussion of the statistical methods used can be found in [39,40].

5 Summary

We have described various technical details of a novel analysis framework introduced in [1] to measure properties of the newly-discovered bosonic state at ~ 125 GeV. This analysis method allows us to fully utilize the power of the golden channel by constructing a continuous likelihood function in all observables, which is also a continuous function of the parameters of interest. We have emphasized in particular the details involved in performing the convolution integral which takes one from 'truth' level observables to detector level observables. We have also briefly discussed other aspects of the framework.

This framework is distinctively different compared to other established methods which utilize templates of discriminants. These template methods offer many advantages such as simplicity and faster analysis set up time, but are more dependent on availability of Monte-Carlo samples and the choice of discriminant. Our framework on the other hand, though more complex and requiring (arguably) more computing resources *prior to analyzing data*, offers great speed, flexibility, and a more complete picture on the extracted results when actually performing the parameter extraction on data. Complete validations of the framework can be found in [39, 40]. With this framework one can then go on to perform a variety of multi-parameter extractions in the $h \to 4\ell$ channel with data obtained at the LHC and other future colliders.

Acknowledgements

We thank the ATLAS and CMS collaborations for their encouragement and interest in this work. The work of R.V.M. is supported by the ERC Advanced Grant Higg@LHC and the Fermilab Graduate Student Fellowship in Theoretical Physics. Y.C. is supported by the Weston Havens Foundation and DOE grant No. DE-FG02-92-ER-40701. Fermilab is operated by Fermi Research Alliance, LLC, under Contract No. DE-AC02-07CH11359 with the United States Department of Energy. This work is also sponsored in part by the DOE grant No. DE-FG02-91ER40684. R.V.M is also grateful to the CalTech physics department for their hospitality during which much of this work was done. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575.

References

[1] Y. Chen, E. Di Marco, J. Lykken, M. Spiropulu, R. Vega-Morales, et al., *8D Likelihood Effective Higgs Couplings Extraction Framework in the Golden Channel*, `arXiv:1401.2077`.

[2] C. A. Nelson, *Correlation Between Decay Planes in Higgs Boson Decays Into W Pair (Into Z Pair)*, *Phys.Rev.* **D37** (1988) 1220.

[3] A. Soni and R. Xu, *Probing CP violation via Higgs decays to four leptons*, *Phys.Rev.* **D48** (1993) 5259–5263, [`hep-ph/9301225`].

[4] D. Chang, W.-Y. Keung, and I. Phillips, *CP odd correlation in the decay of neutral Higgs boson into Z Z, W+ W-, or t anti-t*, *Phys.Rev.* **D48** (1993) 3225–3234, [`hep-ph/9303226`].

[5] V. D. Barger, K.-m. Cheung, A. Djouadi, B. A. Kniehl, and P. Zerwas, *Higgs bosons: Intermediate mass range at e+ e- colliders*, *Phys.Rev.* **D49** (1994) 79–90, [`hep-ph/9306270`].

[6] T. Arens and L. Sehgal, *Energy spectra and energy correlations in the decay $h \to 4\mu$*, *Z.Phys.* **C66** (1995) 89–94, [hep-ph/9409396].

[7] S. Choi, . Miller, D.J., M. Muhlleitner, and P. Zerwas, *Identifying the Higgs spin and parity in decays to Z pairs*, *Phys.Lett.* **B553** (2003) 61–71, [hep-ph/0210077].

[8] C. Buszello, I. Fleck, P. Marquard, and J. van der Bij, *Prospective analysis of spin- and CP-sensitive variables in $h \to ZZ \to 4\ell$ at the LHC*, *Eur.Phys.J.* **C32** (2004) 209–219, [hep-ph/0212396].

[9] R. M. Godbole, . Miller, D.J., and M. M. Muhlleitner, *Aspects of CP violation in the H ZZ coupling at the LHC*, *JHEP* **0712** (2007) 031, [arXiv:0708.0458].

[10] V. Kovalchuk, *Model-independent analysis of CP violation effects in decays of the Higgs boson into a pair of the W and Z bosons*, *J.Exp.Theor.Phys.* **107** (2008) 774–786.

[11] Q.-H. Cao, C. Jackson, W.-Y. Keung, I. Low, and J. Shu, *The Higgs Mechanism and Loop-induced Decays of a Scalar into Two Z Bosons*, *Phys.Rev.* **D81** (2010) 015010, [arXiv:0911.3398].

[12] Y. Gao, A. V. Gritsan, Z. Guo, K. Melnikov, M. Schulze, et al., *Spin determination of single-produced resonances at hadron colliders*, *Phys.Rev.* **D81** (2010) 075022, [arXiv:1001.3396].

[13] A. De Rujula, J. Lykken, M. Pierini, C. Rogan, and M. Spiropulu, *Higgs look-alikes at the LHC*, *Phys.Rev.* **D82** (2010) 013003, [arXiv:1001.5300].

[14] J. S. Gainer, K. Kumar, I. Low, and R. Vega-Morales, *Improving the sensitivity of Higgs boson searches in the golden channel*, *JHEP* **1111** (2011) 027, [arXiv:1108.2274].

[15] B. Coleppa, K. Kumar, and H. E. Logan, *Can the 126 GeV boson be a pseudoscalar?*, arXiv:1208.2692.

[16] S. Bolognesi, Y. Gao, A. V. Gritsan, K. Melnikov, M. Schulze, et al., *On the spin and parity of a single-produced resonance at the LHC*, arXiv:1208.4018.

[17] D. Stolarski and R. Vega-Morales, *Directly Measuring the Tensor Structure of the Scalar Coupling to Gauge Bosons*, *Phys.Rev.* **D86** (2012) 117504, [arXiv:1208.4840].

[18] Y. Chen, N. Tran, and R. Vega-Morales, *Scrutinizing the Higgs Signal and Background in the $2e2\mu$ Golden Channel*, *JHEP* **1301** (2013) 182, [arXiv:1211.1959].

[19] R. Boughezal, T. J. LeCompte, and F. Petriello, *Single-variable asymmetries for measuring the 'Higgs' boson spin and CP properties*, arXiv:1208.4311.

[20] A. Belyaev, N. D. Christensen, and A. Pukhov, *CalcHEP 3.4 for collider physics within and beyond the Standard Model*, arXiv:1207.6082.

[21] P. Avery, D. Bourilkov, M. Chen, T. Cheng, A. Drozdetskiy, et al., *Precision Studies of the Higgs Golden Channel $h \to 4\ell$. Part I. Kinematic discriminants from leading order matrix elements*, arXiv:1210.0896.

[22] J. M. Campbell, W. T. Giele, and C. Williams, *Extending the Matrix Element Method to Next-to-Leading Order*, arXiv:1205.3434.

[23] J. M. Campbell, W. T. Giele, and C. Williams, *The Matrix Element Method at Next-to-Leading Order*, `arXiv:1204.4424`.

[24] **CMS Collaboration** Collaboration, S. Chatrchyan et al., *Search for a Higgs boson in the decay channel H to ZZ(*) to q qbar ℓ^- l+ in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **1204** (2012) 036, [`arXiv:1202.1416`].

[25] **CMS Collaboration** Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys.Lett.* **B716** (2012) 30–61, [`arXiv:1207.7235`].

[26] **CMS Collaboration** Collaboration, S. Chatrchyan et al., *Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs*, *Phys.Rev.Lett.* **110** (2013) 081803, [`arXiv:1212.6639`].

[27] T. Modak, D. Sahoo, R. Sinha, and H.-Y. Cheng, *Inferring the nature of the boson at 125-126 GeV*, `arXiv:1301.5404`.

[28] Y. Sun, X.-F. Wang, and D.-N. Gao, *CP mixed property of the Higgs-like particle in the decay channel* $h \to ZZ^* \to 4l$, `arXiv:1309.4171`.

[29] J. S. Gainer, J. Lykken, K. T. Matchev, S. Mrenna, and M. Park, *Spherical Parametrization of the Higgs Boson Candidate*, *Phys.Rev.Lett.* **111** (2013) 041801, [`arXiv:1304.4936`].

[30] P. Artoisenet, P. de Aquino, F. Demartin, R. Frederix, S. Frixione, et al., *A framework for Higgs characterisation*, `arXiv:1306.6464`.

[31] I. Anderson, S. Bolognesi, F. Caola, Y. Gao, A. V. Gritsan, et al., *Constraining anomalous HVV interactions at proton and lepton colliders*, `arXiv:1309.4819`.

[32] M. Chen, T. Cheng, J. S. Gainer, A. Korytov, K. T. Matchev, et al., *The role of interference in unraveling the ZZ-couplings of the newly discovered boson at the LHC*, `arXiv:1310.1397`.

[33] G. Buchalla, O. Cata, and G. D'Ambrosio, *Nonstandard Higgs Couplings from Angular Distributions in* $h \to Z\ell^+\ell^-$, `arXiv:1310.2574`.

[34] Y. Chen and R. Vega-Morales, *Extracting Effective Higgs Couplings in the Golden Channel*, *JHEP* **1404** (2014) 057, [`arXiv:1310.2893`].

[35] J. S. Gainer, J. Lykken, K. T. Matchev, S. Mrenna, and M. Park, *Beyond Geolocating: Constraining Higher Dimensional Operators in* $H \to 4\ell$ *with Off-Shell Production and More*, `arXiv:1403.4951`.

[36] Y. Chen, R. Harnik, and R. Vega-Morales, *Probing the Higgs Couplings to Photons in* $h \to 4\ell$ *at the LHC*, `arXiv:1404.1336`.

[37] Y.Gao, *talk given at ATLAS Higgs WG Workshop, December 2013*, `https://indico.cern.ch/event/280998/session/5/contribution/29/material/slides/0.pdf`.

[38] R. Vega-Morales, *The Higgs Boson as a Window to Beyond the Standard Model*, `Thesis in fulfillment of P.h.D requirements, Northwestern University (2013)`.

[39] **CMS Collaboration** Collaboration, *Constraints on anomalous HVV interactions using H to 4l decays*, Tech. Rep. CMS-PAS-HIG-14-014, CERN, Geneva, 2014.

[40] Y. Chen, *Precision measurement of the $h(125)$ boson discovered in proton-proton collisions at $\sqrt{s} = 7, 8$ TeV with the CMS detector at the LHC*, `Thesis in preparation for P.h.D. at California Institute of Technology (2014)`.

[41] CMS, *CMS Physics (2006): Technical Design Report Volume 1: Detector Performance and Software*, `https://cds.cern.ch/record/922757/export/hx?ln=en`.

[42] CMS, *CMS Detector Electron Performance*, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/EGMElectronsMoriond2013`.

[43] B. Bradie, *A friendly introduction to numerical analysis: with C and MATLAB materials on website; int. ed.* Prentice-Hall, Englewood Cliffs, NJ, 2006.

[44] F. James, *MINUIT Function Minimization and Error Analysis*, `Reference Manual Version 94.1 (1994)`.