Optimizing Higgs Identification at the LHC

Seung J. Lee
*Department of Physics, Korea Advanced Institute of Science and Technology, 335 Gwahak-ro,
Yuseong-gu, Daejeon 305-701, Korea*

Template overlaps are a class of infrared safe jet observables, based on functional comparison of the energy flow in data with the flow in selected sets (the templates) of partonic states. In a recent work with Leandro G. Almeida, Ozan Erdoğan, José Juknevich, Gilad Perez, and George Sterman, we have shown that template overlap offers the promise of a successful boosted Higgs tagger, have demonstrated how the inclusion of three particle templates allows us to test the influence of gluon emission and color flow, through their effect on energy flow, and have illustrated its use through the construction of partonic template observables.

1 Introduction

The search for the Higgs boson at the Large Hadron Collider is now in the discovery phase with the di-photon signal hinting the existence of ~ 125 GeV Higgs-like particle, and it becomes important to really make sure whether this is indeed the Higgs particle of the Standard Model (SM). It is also very interesting to figure out whether the decay branching ratio of the could-be-Higgs boson is what we expect from the SM, specially in the absence of discovery of any signal for New Physics beyond the SM at the current experimental stage. While $H \to b\bar{b}$ channel is the dominant decay mode for such mass range of the Higgs, the detector sensitivity of this channel is much less than other decay channels due to the swamping QCD background. A jet substructure method was first employed to provide a useful handle to distinguish the signal from the QCD background for the case of boosted Higgs [1] (which is often referred as "BDRS" method). We apply the template method to the same production of a Higgs boson in association with a W boson, $p+p \to W+H$, followed by the dominant light Higgs boson decay, to two b-tagged jets, including schemes for generating templates and for discretizing the data. We investigate the tagging efficiency for this process, and the fake rates from the background process $p+p \to W+jets$.

2 The Template Overlap Method

A template overlap method was developed for the quantitative comparison of the energy flow of observed jets at high-p_T with the flow from selected sets (the templates) of partonic states [2], which can be summarized as follows[2,3]. We denote by $|j\rangle$ the set of particles or calorimeter towers that make up a jet, identified by some algorithm, and take $|f\rangle$ to represent a set of partonic momenta $p_1 \ldots p_n$ that represent a boosted decay, found by the same algorithm. We introduce a functional measure $\mathcal{F}(j,f) \equiv \langle f|j\rangle$ that quantifies how well the energy flow of $|j\rangle$ matches $|f\rangle$. In practice, we find good results with a simple construction of functional overlap based on a Gaussian in energy differences within angular regions surrounding the template partons. Any region of partonic phase space for the boosted decays, $\{f\}$, defines a template. We use our knowledge of the signal and background to design a custom analysis for each resonance, to make use of differences in energy flow between signal and background. We define the template overlap of observed jet j as $Ov(j, f[j]) = \max_{\{f\}} \mathcal{F}(j, f)$, the maximum functional overlap of j to a state $f[j]$ within the template region. We use the notation $f[j]$ for the state of maximum overlap to emphasize that the value of the overlap functional depends not only on the physical state $|j\rangle$, but also on the choice for the set of template functions f.

As a simple working example, for an N-particle final state, we represent our template overlap (dropping the superscript (F)) as [2]

$$Ov_N(j, p_1, \ldots, p_N) = \max_{\tau_N^{(R)}} \exp\left[-\sum_{a=1}^{N} \frac{1}{2\sigma_a^2} \left(\int d^2\hat{n}\, \frac{dE(j)}{d^2\hat{n}} \theta_N(\hat{n}, \hat{n}_a^{(f)}) - E_a^{(f)} \right)^2 \right], \tag{1}$$

where the direction of template particle a is $\hat{n}_a$ and its energy is $E_a^{(f)}$. In applications below, we will use these energies to set the widths of the Gaussians. The functions $\theta_N(\hat{n}, \hat{n}_a^{(f)})$ restrict the angular integrals to (nonintersecting) regions surrounding each of the template momenta. We will refer to the corresponding state as the "peak template" $f[j]$ for state j. The peak template $f[j]$ provides us with potentially valuable information on energy flow in j. The output of the peak template method for any physical state j is the value of the overlap, $Ov(j, f)$, and also the identity of the template state $f[j]$ to which the best match is found.

2.1 Construction of template functions for the Higgs

Here we have adopted the expectation that a good rejection power is obtained when we simply use the signal (Higgs) distribution itself to construct our templates[2]. Then, we want our template overlaps to be functionals of the energy flow of any specific event (usually involving jets), which we label j, and a model, or template, for the energy flow in a signal, referred to as f . The number of particles in the templates is not necessarily fixed, and templates with more than the minimum number of particles are possible. We find, however, that combining templates in the full phase space for $N = 3$ and $N = 2$ already delivers encouraging results for the Higgs [3]. Our templates will be a set of discretized partonic states corresponding to given points in phase space. We generate a large number of template states enough to sufficiently cover both two- and three-particle phase space for Higgs decay.

2.2 Selection and Discretization of the Data

We generate events for $W^+ + H \to l^+\nu_l\, b\bar{b}$ and $W^+ + jets \to l^+\nu_l + jets$ in a configuration with large transverse momentum, using PYTHIA 8.150[4], SHERPA 1.3.0[5] (with CKKW matching [6]), and MADGRAPH[7] interfaced to PYTHIA 6 [8] (with MLM matching[9]). Jets are reconstructed using FASTJET [10], and the anti-k_T algorithm [11] with large effective cone size $R = 0.7$. We have chosen plausible value for R, based on a combination of physics input and a trial-and-error, but have not attempted to optimize them systematically. For each event, we find the jet with the highest transverse momentum and impose a jet mass window for the Higgs. We choose the jet mass window to be 110 GeV $\leq m_J \leq$ 130 GeV, with our reference Higgs boson mass chosen to be $m_H = 120$ GeV, and jet energy 950 GeV $\leq P_0 \leq$ 1050 GeV. This gives us a set of final states j.

We compute the overlap between data state j and two- or three-body template f from the unweighted sum of the energy in the nine cells of state j surrounding and including the occupied cells of template state f[2],

$$Ov_N(j,f) = \max_{\tau_N^{(R)}} \exp\left[- \sum_{a=1}^{N} \frac{1}{2\sigma_a^2} \left(\sum_{k=i_a-1}^{i_a+1} \sum_{l=j_a-1}^{j_a+1} E(k,l) - E(i_a,j_a)^{(f)} \right)^2 \right], \qquad (2)$$

where $N = 2$ or 3. Here, $E(i_a, j_a)^{(f)}$ is the energy in the template state for particle a whose direction is labelled by indices i_a and j_a. If one of the sums extends outside the jet cone, we set the corresponding energies $E(k,l)$ to zero. We fix σ_a (for the ath parton) by that parton's energy, $\sigma_a = E(i_a, j_a)^{(f)}/2$.

3 Summary of template overlaps for Higgs and QCD jets

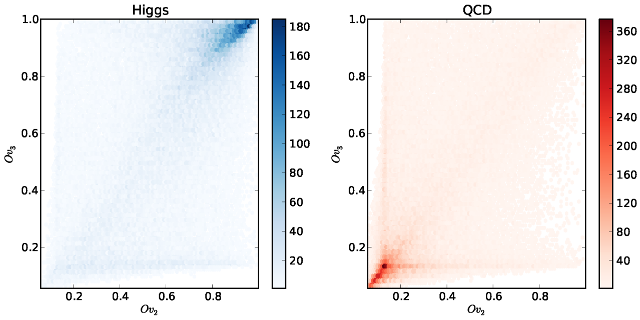

Figure 1: Density plots of 2-body overlap *vs.* 3-body overlap for boosted Higgs and QCD jets with $R = 0.7$.

Both two- and thee-body template overlap have substantial discriminating power, which can be seen in the scatter plots, shown in Fig. 1, of Ov_2 and Ov_3 for Higgs signal (left panel) and dijet production (right panel)[3]. While the signal events cluster around the upper right corner of the plot, most QCD jet events are localized diagonally opposite in the lower left. It follows immediately that making tight cuts on each observable, by drawing a rectangular window in the upper right corner of the scatter plot, makes a good discriminator to separate signal from background.

Table 1: Efficiencies and fake rates for jets with $R = 0.7$ (using anti-k_T: $D = 0.7$), 950 GeV$\leq P_0 \leq$1050 GeV, 110 GeV$\leq m_J \leq$130 GeV and $m_H = 120$ GeV. We have imposed various cuts on Ov, $\bar{\theta}$ and Pf variables: f$Ov_2 > 0.8$, $Ov_3 > 0.8$, $\bar{\theta} < 0.4$ and $Pf < 0.2$ (for Sherpa, we had $Ov_3 > 0.7$, $\bar{\theta} < 0.45$ and $Pf < 0.3$ for comparable results).

MC	Jet mass cut only		Mass cut + Ov ($+\bar{\theta}+$ Pf)	
	Higgs-jet efficiency [%]	fake rate [%]	Higgs-jet efficiency [%]	fake rate [%]
PYTHIA 8	70	10	10	0.05
MG/ME	70	10	10	0.05
SHERPA	60	10	10	0.05

The final results for the Higgs jet case are summarized in Tables 1[3] for the three event generators and $R = 0.7$, that result from including these simple, naive one-dimensional cuts in Ov_2, Ov_3 (with the three-body angular variable vaiable $\bar{\theta}$[a] and planar flow Pf[12,13]) at fixed

---

[a] The three-body angular variable $\bar{\theta}$ is defined as $\bar{\theta} = \sum_i \sin\theta_{iJ}$ where θ_{iJ} is the angles between the jet axis and the template momenta. It is a partonic level variable, which becomes a physical observable when it is associated with the peak template.

signal efficiency of $S = 10\%$. We find a large enhancement of signal compared to background, typically of the order of fifteen or more. Taking into account the rejection of QCD jets by imposing a mass window, these numbers (for a single massive jet) are multiplied by factors of ten to twenty.

The template-based approach yields, without optimization, moderately improved numbers compared with those found from other methods in the literature (see for example Ref. [1,14]). Our template overlap method has an advantage for dealing with pile-up issue, as it is based on parton-hadron dulaity where the spikiness of the jet energy distribution naturally avoids the complication of pile-up issue, as well as it provides a method based on first principle. The template overlap method is quite general, and it can be used for other massive object searches as well. Note that our event selection was chosen in a kinematical regime that at present is unrealistic for the LHC. However, our findings should serve as a proof of concept for many of the ideas, and, based on ongoing research[15,16], we expect an extended phenomenological analysis to deliver similar qualitative behaviour in terms of rejection power. As the LHC continues to explore the energy frontier of particle physics, template overlap provides us with an interesting tool for further development of jet substructure techniques.

References

1. J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].
2. L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung, Phys. Rev. **D82**, 054034 (2010). [arXiv:1006.2035 [hep-ph]].
3. L. G. Almeida, O. Erdogan, J. Juknevich, S. J. Lee, G. Perez and G. Sterman, arXiv:1112.1957 [hep-ph].
4. T. Sjostrand, S. Mrenna and P. Skands, Comput. Phys. Commun. **178**, 852 (2008) [arXiv:0710.3820 [hep-ph]].
5. T. Gleisberg, S. Hoche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, JHEP **0902**, 007 (2009) [arXiv:0811.4622 [hep-ph]].
6. S. Hoeche, F. Krauss, S. Schumann and F. Siegert, JHEP **0905**, 053 (2009) [arXiv:0903.1219 [hep-ph]].
7. F. Maltoni and T. Stelzer, JHEP **0302**, 027 (2003) [arXiv:hep-ph/0208156]; T. Stelzer and W. F. Long, Comput. Phys. Commun. **81**, 357 (1994) [arXiv:hep-ph/9401258]. JHEP **0709**, 028 (2007) [arXiv:0706.2334 [hep-ph]].
8. T. Sjostrand, S. Mrenna and P. Skands, JHEP **0605**, 026 (2006) [arXiv:hep-ph/0603175].
9. M. L. Mangano, M. Moretti, F. Piccinini and M. Treccani, JHEP **0701**, 013 (2007) [arXiv:hep-ph/0611129].
10. M. Cacciari and G.P. Salam, Phys. Lett. B 641 (2006) 57 [hep-ph/0512210]; M. Cacciari, G.P. Salam and G. Soyez, http://fastjet.fr/
11. M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804**, 063 (2008) [arXiv:0802.1189 [hep-ph]].
12. J. Thaler and L. T. Wang, JHEP **0807**, 092 (2008) [arXiv:0806.0023 [hep-ph]].
13. L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung and J. Virzi, Phys. Rev. D **79**, 074017 (2009) [arXiv:0807.0234 [hep-ph]].
14. D. E. Soper and M. Spannowsky, Phys. Rev. D **84**, 074002 (2011) [arXiv:1102.3480 [hep-ph]].
15. M. Backovic, J. Juknevich, G. Perez and J. Winter, In preparation.
16. G. Gur-Ari, M. Field, D.A. Kosower, L. Manelli and G. Perez, In preparation.