POLYMORPHIC GRAM-SCHMIDT LINEAR-PADDING IMPLEMENTATION OF NUISANCE PARAMETERS

Mihai-Tiberiu DIMA [1*]

Nuisance parameters are used to model unexpected data features beyond ones already contained in a model. Without their contribution, data upsetting features distort the fit, and lead to erroneous results. Nuisance parameters model the extra features in the data that were not anticipated, in an ad-hoc manner and are consequently difficult to determine. Further more, their contribution may overlap with the model's signal and thus affect precisely the measurement. I present here a general polynomial method that addresses the issue. With a set of orthogonal polynomials I "pad" the difference between data and optimal signal-plus background fit (both of known, fixed shape). This eliminates the pressure on the latter feel from the fit's minimisation, which forces unrealistic coefficients for the signal. Study case for the method is a fit to K^{*0} production at very high momentum in e^+e^- collisions at the Z^0 energy. Using electron beam polarisation this measurement was used for the direct determination of quantum chromodynamics strangeness suppression by the SLD Collaboration.

Keywords: Particle Physics, Numerical Methods.

1. Introduction

High energy physics experiments model data with the aim of unravelling fundamental laws of nature, in particle collisions at very high energies. The process of extracting meaningful physics results from the experimental data is complicated by the presence of data deviations from the simplifying assumptions of the adopted models. Ad-hoc modelling of these deviations with "nuisance parameters" can phenomenologically describe the data in detail [4]. These may represent uncertainties, or systematic effects that are not of primary interest, and which could otherwise impact the final measurement. Examples in high energy physics of nuisance-parameter augmented models are in:

- *Detector uncertainty*: such as calibration, efficiency, and resolution, which can distort the measured properties of particles. For instance, the precise energy measurement of particles in a calorimeter can be affected by variations in the detector response, introducing a systematic uncertainty in the final result. A number of Higgs related searches performed at ATLAS

* Corresponding author
[1] UPB FSA, Romania, e-mail: mihai_tiberiu.dima@stud.fsa.upb.ro

(CERN/LHC) modeled detector-related uncertainties through nuisance parameters [5].

- *Theoretical uncertainties*: limitations in the current understanding of fundamental physics can lead to uncertainties in theoretical predictions. For example, uncertainties in the values of fundamental constants, or the precise form of theoretical models used to describe particle interactions can affect the interpretation of experimental data. An example for the top quark production cross section, is in [6].
- Machine Luminosity Modeling - for example in determining the luminosity of the machine during a given run period. This is crucial in predicting the number of signal and background events, thereby affecting the final measurement. A good example is found in [7].
- Pileup Uncertainty - multiple particle collisions within the same bunch crossing can overlap with the interaction of interest, obstructing the isolation of the latter. In the measurement of Higgs rare decays for instance, pileup increases significantly background and complicates the analysis. A relevant example is in [8].
- Statistical uncertainties - where data features and detector bias are modeled with extra phenomenologically determined parameters. A comprehensive overview of statistical methods used in high energy physics, together with techniques for handling nuisance parameters can be found in the F. James books [2]. A practical example is in background modelling, where unwanted events mimic the signal of interest and can significantly bias the measurement if they are not properly accounted for.

Finding proper nuisance parameters can thus pose a significant challenge in high energy physics analyses, as in their absence unwanted effects may obscure the true signal, or bias it in non-deconvolutable ways. It is therefore crucial to develop robust methods to determine these parameters and ensure the accuracy and reliability of the experimental results.

2. K^{*0} production fit

When data is statistically challenging, it is usual to fix contributor shapes to previously determined models, and let solely the amplitudes vary in the fit. To exemplify this I approached an interesting measurement of strangeness suppression, that used K^{*0} decays in the SLD detector [1] at the Stanford Linear Accelerator Center. The philosophy of the measurement is the following: in e^+e^- collisions at $\sqrt{s} = 91.2GeV$, very high momentum K^{*0}'s are assumed to contain the primary $\bar{s}$ antiquark (moving in the hemisphere of the e^- for right-handed beam polarisation), while those moving in the same direction for left-handed

beam polarisation are assumed to contain the primary d-quark. This is the leading-particle hypothesis, that very high momentum particles from the hadronisation process must contain one of the primary quarks. Now, it is harder for the light d-quarks to pull out of QCD (Quantum Chromodynamics) vacuum the heavier $\bar{s}$ antiquark, than vice-versa. The ratio is the strangeness suppression parameter γ_s. The SLD experiment having benefited of above 75% e^- beam polarisation [1] could infer the quark and antiquark hemispheres through electroweak forward-backward asymmetry. This allowed for a model-independent determination of strangeness suppression, unlike numerous other measurements, that relied heavily on hadronisation models for the measurement.

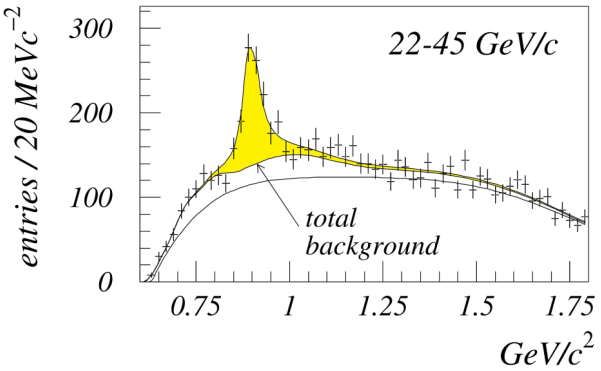

Fig. 1. $K\pi$ invariant mass showing the K^{*0} fit in the 22-45 GeV/c^2 energy range as performed by the SLD experiment [3]. The shape of the components for the fit were determined, in part by Monte Carlo (figure 2), and for the signal by a relativistic Breit-Wigner distribution. The total background is indicated by the arrow, while the purely random-combinatorial one is the line below it.

Figure 1 shows the $K\pi$ invariant mass for a K^{*0} production fit in the 22-45 GeV/c^2 energy range as performed by the SLD experiment [3]. In this case the statistics seems minimally sufficing and the fit looks good.

As it will become evident, this is a good prototype case for using nuisance parameters in the fit. This is because the distribution contains a number of reflections, the most significant outlined in figure 2. Their shapes are determined from Monte Carlo, and their amplitudes from the fit in figure 1.

Now, let's divide the data, as outlined by the philosophy of the measurement, in two parts - shown in figure 3: the "*enhanced*" sample, consisting of K^{*0} candidates going in the antiquark-hemisphere (direction of e^- for right-handed beam polarisation) and the "*depleted*" sample, consisting of K^{*0} candidates going in the quark-hemisphere (direction of e^- for left-handed beam polarisation). For statistical enrichment, each sub-sample is united with its mirror-case, as shown in the figure. The ratio of the production rates of the two is γ_s. Of

course, the real-picture includes other influences such as K^{*0} production from b and c quark decays, the imperfection of quark-hemisphere tagging, purity of *uds* light-quark event selection, etc. These are accounted for in the subsequent unfolding procedure.

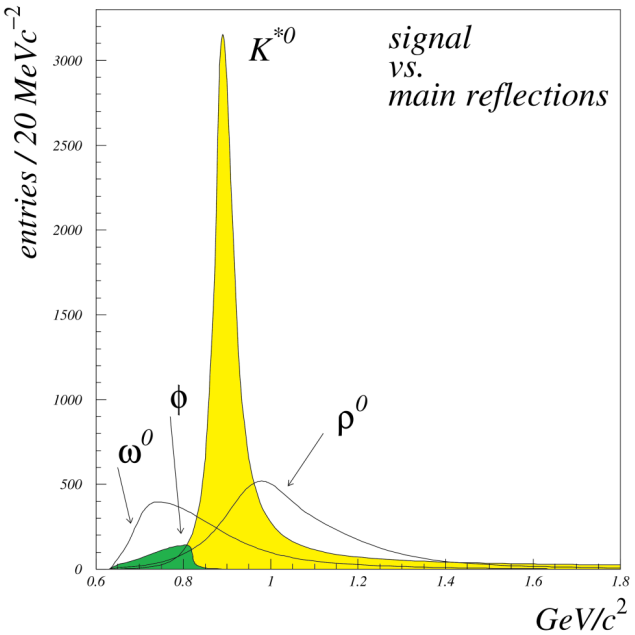

Fig. 2. Shapes of various Monte Carlo components in a $K\pi$ invariant mass plot, together with the relativistic Breit-Wigner K^{*0} peak (not to scale) - SLD experiment [3]. Note the $\omega^0 - \rho^0$ saddle shape and the φ shoulder to the left of the signal, all posing problems in the fit.

From the fit's point of view, there is not enough statistics for a precision fit, where I let all parameters vary (without shape distortion risk). Therefore I fix the shapes of background and signal and then perform a fit for the amplitudes of the two.

This is all good when the data aligns with the model, however if there are upsetting aspects in the data (in this case the small peak in the 0.8 GeV/c^2 region in figure 3), the fit will not work perfectly. It will try to adjust the amplitudes of the two shapes it has to best fit all data. This distorts the signal, and we are of course interested in correcting this.

To this end I introduced a polynomial-"padding" for the new shape in the 0.8 GeV/c^2 region. Typically, this is done with ad-hoc parameterizations, however such also have overlaps with the signal and affect the measurement. As can be noted, the problem is not as trivial as it seems at first glance.

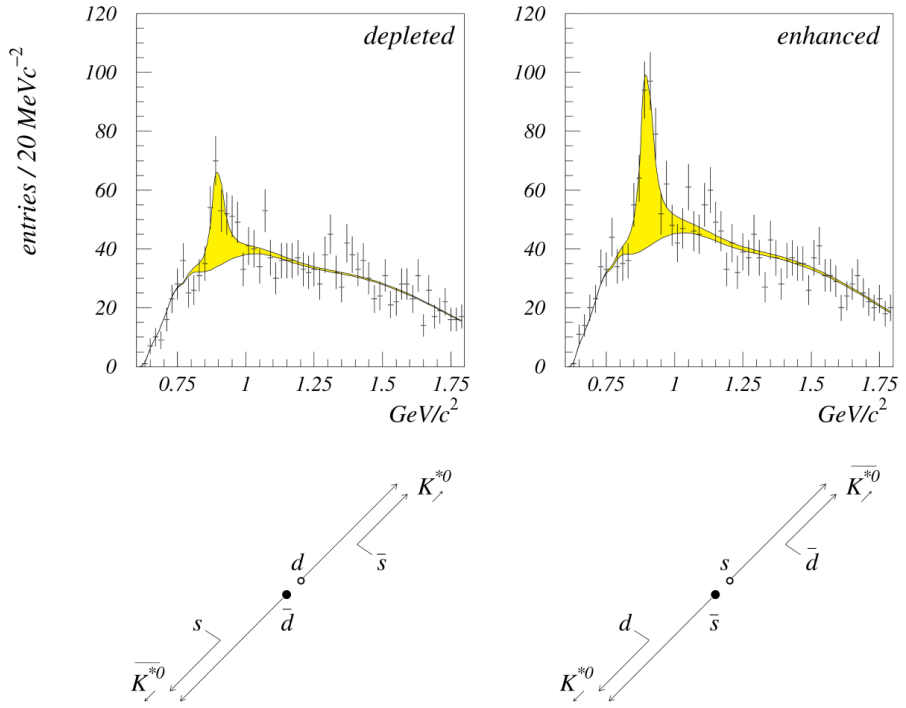

Fig. 3. Original [3] fits to the "enhanced" (K^{*0} → antiquark, right-handed e−) and "depleted" (K^{*0} → quark, left-handed e⁻)" samples. The upper left insert features the fit to the "depleted" sample, with the lower left insert showing the two cases comprised. The upper right insert features the fit to the "enhanced" sample, with the lower right insert showing the two cases comprised.

3. Polynomial padding

It is important the added padding to have a mathematical basis. Since there is no model for the padding, I formulated indirect ones: after the fit is done, the residuals should look like "noise". However I do not know how noise looks like - and thus I defined noise as not composed of "known" shapes.

The padding will be a sum of terms:

$$pad(x) = C_a adhoc_a(x) + C_b adhoc_b(x) + \ldots + \sum_{k=1}^{n} C_k P_k(x). \quad (1)$$

where the adhoc terms are a set of (optional) preferred shapes and the polynomials $P_k(x)$ are an orthonormal set over the interval considered for the fit. This can be re-written as:

$$pad(x) = [C_a + C_b\,(adhoc_a|adhoc_b)]adhoc_a(x) + C_b adhoc_b^{\perp}(x) + \ldots + \sum_{k=1}^{n} C_k P_k(x). \quad (2)$$

I could have equally well assumed from the begining that all terms in the sum are orthonormal (obtainable through the Gram-Schmidt orthogonalisation method [9]). This answers also the question about what are "known" shapes, respectively, the terms in the above sum. This makes sense: I chose a set of terms to model all that is inconvenient - thus all else is "non-inconvenient noise".

The conditions then become:

$$(data\ - C_S\ \cdot sgn - C_B \cdot bckg - pad|known\ shapes) = 0. \tag{3}$$

respectively, I extracted out all "known" shapes, therefore the rest is "non-inconvenient noise". This means the fit becomes:

$$\|data_\perp\ - C_S\ \cdot sgn_\perp - C_B \cdot bckg_\perp\|^2 = min. \tag{4}$$

where:

$$\begin{aligned} data_\perp(x) &= data(x) - adhoc_a(x) - (adhoc_a|data) \\ &-\ldots-\sum_{k=1}^{n} P_k(x)\cdot(P_k|data), \\ sgn_\perp(x) &= sgn(x) - adhoc_a(x) - (adhoc_a|sgn) \\ &-\ldots-\sum_{k=1}^{n} P_k(x)\cdot(P_k|sgn), \\ bckg_\perp(x) &= bckg(x) - adhoc_a(x) - (adhoc_a|bckg) \\ &-\ldots-\sum_{k=1}^{n} P_k(x)\cdot(P_k|bckg). \end{aligned} \tag{5}$$

i.e. - I extracted out of the above everything that looks like adhoc shapes (none if there are no preferred shapes in use) and like the set of polynomials. Here "$(a|b) = \sum_i \bar{a}_i b_i$" is the scalar product (in discrete form) of the shapes a and b. Therefore the residual can be further minimised only by using $sgn_\perp$ and $bckg_\perp$ (on the remainder-space, on which the polynomial-padding cannot act and minimise).

The $sgn_\perp$ and $bckg_\perp$ coefficients are determined from equation (4) by derivation and equalling to zero, via a 2×2 linear equations system:

$$\begin{aligned} (sgn_\perp|data) &= C_S(sgn_\perp|sgn_\perp) + C_B(sgn_\perp|bckg_\perp), \\ (bckg_\perp|data) &= C_S(bckg_\perp|sgn_\perp) + C_B(bckg_\perp|bckg_\perp), \end{aligned} \tag{6}$$

respectively:

$$C_S = \frac{(bckg_\perp|bckg_\perp)(sgn_\perp|data) - (sgn_\perp|bckg_\perp)(bckg_\perp|data)}{det},$$
$$C_B = \frac{(sgn_\perp|sgn_\perp)(bckg_\perp|data) - (sgn_\perp|bckg_\perp)(sgn_\perp|data)}{det}, \quad (7)$$

where $det = (sgn_\perp|sgn_\perp)(bckg_\perp|bckg_\perp) - (sgn_\perp|bckg_\perp)(bckg_\perp|sgn_\perp)$.

For numerical reasons the procedure is performed in 2 steps: I use $(C_S + \Delta C_S)sgn(x) + (C_B + \Delta C_B)bckg(x)$as final result and in the first step determine (C_S, C_B) from:

$$\|data\ - C_S\ \cdot sgn - C_B \cdot bckg\|^2 = min. \quad (8)$$

Subsequently I define $data'(x) = data(x) - C_S\ \cdot sgn(x) - C_B \cdot bckg(x)$ and apply the procedure described above to $data'(x)$ to obtain $(\Delta C_S, \Delta C_B)$, which I sum with (C_S, C_B) to obtain the final answer.

4. Example result

I applied the padding procedure to the data in figure 3, with 2 backgrounds (a random combinatorial background and one due to "reflections" - as shown in figure 2). This follows the same recipe, just that the final linear system is 3×3, with $sgn_\perp(x)$, $rcbg_\perp(x)$ and $refl_\perp(x)$.

The plots are shown in figure 4. Note that the resonance-like shape on the left has been accounted for. Overall, the signal was reduced by 13.9% (depleted) and 10.4% (enhanced) showing that in both cases the signal-shape was trying previously to compensate for the extra features in the data.

The final answer cannot be infered from these figures however, as the unfolding procedure requires the full nuisance-programme applied to the Monte Carlo data and to the analysis of the contamination from heavy-flavor decays (contaminations which change with how the fit accounts for the various shapes).

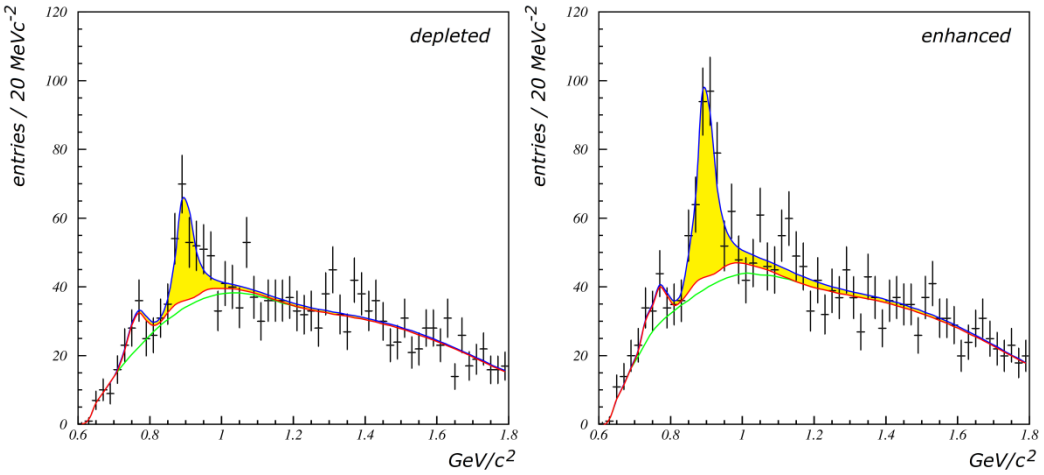

Fig. 4. Fits with the polynomial-padding method. The signal shows a 13.9% (depleted) respectively and 10.4% (enhanced) reduction, following the adjustment the polynomialpadding performed. This shows that previously the fit was overcompensating. While the ratio changes by only 3.9%, the mechanics of my method is well evidentiated.

This borders beyond the scope of the current paper, and is significantly more involved, however a ~3-4% decrease in γ_s below the already lower-side value reported in [3] (0.225 ± 0.068 ± 0.011) could likely be expected. This reinforces the idea that strangeness production in quark-gluon plasma environments is higher than in pure $q\bar{q}$ hadronisation of primary-quarks from e+e- collisions - possibly due to the larger color field fluctuations.

5. PLN polymorphic C++ polynomials class

The software for this method was implemented via a polymorphic C++ class for polynomials. There is ample debate into how polymorphism should be implemented in C++. There is the P1240 proposal for scalable reflection (Andrew Sutton et al.) [10] based on static value evaluation (i.e. - at compile time). Then there is the BOOST library, which is rather heavy syntax and slow due to the use of containers. There is Andrei Alexandrescu's ObjectFactory [11], but this is yet more CPU intensive. Lastly there is C++ 17 `std :: variant`, `std :: visit` which are completely untenable syntax wise. I adopted a simplified template machanism, paralleling in a sense P1240. The main point is that in science we don't use that many types, and for the ones existing we can mimick run-time polymorphism with a (SFINAE based) static type calculator, that performs calculations like:

```
cpx<int> *pln<double> = pln<cpx<double>>        (9)
```

The downside is that I had to instantiate the order of 1280 operators and functions for all type combinations. Rather tedious, but in the end do-able (I placed all instantiations in a file type ⋆.ie where I followed simple rules, such as to list all possible combinations). The advantage is that each operator is now a fast operator. There are no V-tables, no internal type encoding/math to be done.

The polynomials are represented by N-dimensional vectors, where N is user defined. The object resource is move-semantix compliant and R-valued operators are implemented (bringing the total to 7 operator types for each algebraic operation). This covers all practical applications and encompasses any function type (contrary to an algebraic implementation, where a calculator-parser would have been needed). I did implement a mini-parser for printing the polynomials, which takes the names in an expression and "computes" a new name, equal to the algebraic expression of the two.

The code understands scalar product expressions, such as `G = G - (F|G) * F` (for instance for Gram-Schmidt calculations), which is very convenient. This the "*Math-on-Paper*" paradigm, by which packages should allow the user to write code very similar to writing math on paper, making the implementation of the equations in the code very simple. Some examples are:

- *Matrix element*: *(psi| **A** |psi)* within the code will compute the matrix element, which is written the same on paper when doing the equations;
- *Diadic product*: *(a^b)* calculates the diadic product between vectors *a* and *b*, with no other hassle;
- *Frobenius norm*: *fabs(**A**)* will compute the Frobenius norm of matrix A.

6. Conclusions

I presented a novel method for the implementation of nuisance parameters, with the help of a set of ancillary orthogonal polynomials. Study case for the method was the fit for the production of K^{*0} at very high momentum in e^+e^- collisions at the Z^0 energy, following the SLD Collaboration [3]. I estimate the impact of the method to be in the -3-4% range, however this example shows visually very well the purpose of nuisance parameters. The method accounts for the upsetting data feature seamlessly and proves its modus operandi, padding the "uneasy" parts between model and data.

REFERENCES

[1] *K. Abe et al. (SLD Collaboration)*, SLD Design Report, SLAC-PUB-273, 1984; *K. Abe*
[2] *et al. (SLD Collaboration)*, Highlights of the SLD Physics Program at the SLAC Linear
[3] Collider, SLAC-PUB-8985, 2001.
[4] *F. James*, Statistical Methods in Experimental Physics, World Scientific Publishing,
[5] ISBN: 978-981-270-527-2, 2006.

[6] *K. Abe et al. (SLD Collaboration)*, Measurement of Leading Particle Effects in Decays
[7] of Bosons into Light Flavors, Phys. Rev. Lett., **78**(1997), 3442–3446; unpublished SLD
[8] data used with permission.
[9] *L. Demortier*, P-values and nuisance parameters, PhyStat2007: Proceedings of the
[10] Statistical Problems in Particle Physics, Astrophysics, and Cosmology, pp. 23-27, 2007.
[11] *G. Aad et al.*, Observation of a new particle in the search for the Standard Model Higgs
[12] boson with the ATLAS detector at the LHC, Phys. Lett., **B716**(2012), 1-29.
[13] *Tumasyan A. et al. (CMS Collaboration)*, Measurement of the top quark mass using a
[14] profile likelihood approach with the lepton + jets final states in proton–proton collisions
[15] at $\sqrt{s}$ = 13 TeV, European Physical Journal C, **83**(2023), ID 963.
[16] $\sqrt{}$ *Aaboud M. et al. (ATLAS Collaboration)* s = 13 TeV using the ATLAS detector at the
LHC, European Physical Journal C, , Luminosity determination in pp collisions at
[17] **83**(2023), ID 982.
[18] *Tumasyan A. et al. (CMS Collaboration)*, Pileup mitigation at CMS in 13 TeV data,
[19] Journal of Instrumentation, **15**(2020), ID P09018.
[20] *L. Pursell, S.Y. Trimble*, Gram-Schmidt Orthogonalization by Gauss Elimination, The
[21] American Mathematical Monthly, **98**(1991), 544-549.