PROCEEDINGS OF SCIENCE

Combination of highly correlated measurements of the muon spin precession frequency in a magnetic field for the FNAL measurement of the muon magnetic anomaly

Alberto Lusiani[a,b,*] on behalf of the Muon $g-2$ collaboration

[a] *Scuola Normale Superiore, Pisa, Italy*
[b] *INFN, sezione di Pisa, Pisa, Italy*

E-mail: alberto.lusiani@sns.it

We describe the procedures that were developed to verify the consistency and combine the multiple measurements of the muon spin precession frequency by the Muon $g-2$ collaboration for the muon magnetic anomaly measurement reported in 2023. To properly verify the consistency of different analyses up to order of 30 ppb, correlations have been modeled and estimated, exploiting bootstrap techniques. A combination procedure has been designed to combine highly correlated measurements to obtain a robust final result with a conservative sub-ppm uncertainty.

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

*Speaker

1. Introduction

The recent muon magnetic anomaly measurement by the Muon $g-2$ collaboration [1] relies on sub-ppm measurements of the muon spin precession frequency ω_a in a magnetic field, which has been performed on three datasets collected in 2019 and 2020 (Run-2, Run-3a, Run-3b). On each dataset, 7 independent analysis groups completed a total of 19 measurements, using different methods and fit models. Within each dataset, the averaging of the ω_a measurements must be performed by taking into account that they are highly correlated, and that the amount of correlation is only known with limited precision, especially when it involves systematic uncertainties. In the following, we describe the procedures that have been designed and used.

2. Consistency checks and averaging with two measurements

It is well known [2] that when averaging two measurements A and B with uncertainties $\sigma_A < \sigma_B$ with the minimum χ^2 method (identical to the Best Linear Unbiased Estimator (BLUE) method), the weight of the less precise B measurement is:

$$w_B = \frac{\sigma_A^2 \left(1 - \rho(\sigma_B/\sigma_A)\right)}{\sigma_A^2 - 2\rho\sigma_A\sigma_B + \sigma_B^2} \,, \tag{1}$$

which is positive when the correlation $\rho < \sigma_A/\sigma_B$, zero for $\rho = \sigma_A/\sigma_B$ (referred to as "critical" correlation hereafter), and negative for $\rho > \sigma_A/\sigma_B$. When two measurements performed on the same data differ only for their statistical precision, $\rho < \sigma_A/\sigma_B$ and the optimal average weights are both $w_i \geq 0$, as discussed for a specific example in section 7.6.2 of Ref. [2]. The less precise measurement B contributes to the precision of the optimal average, and even in the limit case of $\rho = \sigma_A/\sigma_B$ and $w_B = 0$ its consistency with the more precise measurement is useful to check on procedural errors and systematic uncertainties. The pull between two measurements defined as

$$p_{AB} = \frac{v_B - v_A}{\sigma_{B-A}} \,, \tag{2}$$

and corresponds to the difference or residual between the two measurements values v_A and v_B divided by its uncertainty σ_{B-A}. Assuming Gaussian uncertainties, p_{AB} is expected to be Normally distributed. The residual uncertainty is $\sigma_{B-A}^2 = \sigma_A^2 - 2\rho\sigma_A\sigma_B + \sigma_B^2$ and, for critically correlated measurements, $\sigma_{B-A}^2 = \sigma_B^2 - \sigma_A^2$.

3. Correlations between ω_a measurements

The statistical uncertainties of the 19 ω_a measurements reported in Table 1 are determined by the respective analysis method categories, labelled A (asymmetry), T (threshold) and Q (charge). To a first but good approximation, the A measurements are known to be statistically optimal [3], and it is expected and observed that $\sigma_A < \sigma_T < \sigma_Q$, with A and T critically correlated. Since the Q method analyses differ in several respects from the A and T ones, the expected correlation cannot be reliably estimated and we have assumed that they are also critically correlated to both the A and T measurements, i.e., $\rho_{QX} = \sigma_X/\sigma_Q$, with $X = A, T$. When performing consistency checks, this

Table 1: Run-2, Run-3a and Run-3b dataset ω_a measurements' values and statistical uncertainties, expressed as ppm deviations from a reference ω_a value. The "Group", "Method category", "Ratio", "Reco" columns report measurements' properties that are discussed in the text.

Group	Method category	Ratio	Reco	Run-2 [ppm]	Run-3a [ppm]	Run-3b [ppm]
Gr1	T	Ra1	Re1	-99.147(382)	-98.726(329)	-97.304(528)
Gr2	T	Ra1	Re4	-99.047(378)	-98.581(325)	-97.145(522)
Gr3	T	Ra1	Re3	-99.198(377)	-98.690(323)	-97.267(520)
Gr4	T	Ra1	Re4	-99.029(378)	-98.603(325)	-97.191(513)
Gr5	T	Ra1	Re2	-99.112(377)	-98.682(320)	-97.298(520)
Gr6	T	Ra1	Re1	-99.171(376)	-98.700(323)	-97.274(519)
Gr1	A	Ra1	Re1	-99.199(344)	-98.430(295)	-97.438(476)
Gr2	A	Ra1	Re4	-99.157(340)	-98.397(293)	-97.316(470)
Gr3	A	Ra1	Re3	-99.253(337)	-98.416(291)	-97.422(468)
Gr4	A	Ra1	Re4	-99.134(340)	-98.416(291)	-97.337(466)
Gr5	A	Ra1	Re2	-99.197(339)	-98.355(290)	-97.453(468)
Gr6	A	Ra1	Re1	-99.232(338)	-98.408(290)	-97.407(467)
Gr1	T	Ra3	Re1	-99.160(383)	-98.710(329)	-97.244(529)
Gr3	T	Ra2	Re3	-99.189(383)	-98.693(334)	-97.279(533)
Gr4	T	Ra2	Re4	-99.006(384)	-98.549(324)	-97.158(513)
Gr1	A	Ra3	Re1	-99.180(345)	-98.432(297)	-97.372(477)
Gr3	A	Ra2	Re3	-99.222(345)	-98.458(301)	-97.402(480)
Gr7	Q	Ra1	Re5	-99.191(543)	-98.555(414)	-96.875(663)
Gr7	Q	Ra2	Re5	-99.300(491)	-98.638(386)	-97.239(616)

is a conservative choice, because if the actual correlation is smaller, the residual uncertainty would be larger and the pull would be smaller. Actual correlations larger than the critical threshold are not expected for measurements performed on the same data and differing only for their statistical precision, as already mentioned in Section 2. Analyses differ for a series of features in addition to the method category. Two features, reported on Table 1 are relevant for estimating the correlations: the use of the "Ratio" method, which has three options reported in the table column "Ratio", and the reconstruction of the energy deposited in the calorimeter by a muon-decay positron, which has 5 options reported on the table column "Reco". Further details on the above-mentioned analysis features are not relevant for this document and are reported elsewhere [4]. The ω_a measurements uncertainty is primarily determined by the method category A, T and Q, and analyses that only differ by one or both "Ratio" or "Reco" features have approximately the same total uncertainty, in comparison. We model the covariance between two measurements A and B as:

$$V_{AB} = \begin{pmatrix} \sigma_A^2 & \frac{\sigma_A^2+\sigma_B^2}{2} - \frac{|\sigma_B^2-\sigma_A^2|}{2} - \left(\sum_i \sigma_{r,i}^2\right) \cdot \frac{\sigma_A^2+\sigma_B^2}{2} \\ \frac{\sigma_A^2+\sigma_B^2}{2} - \frac{|\sigma_B^2-\sigma_A^2|}{2} - \left(\sum_i \sigma_{r,i}^2\right) \cdot \frac{\sigma_A^2+\sigma_B^2}{2} & \sigma_B^2 \end{pmatrix}, \quad (3)$$

where the dimensionless relative uncorrelated uncertainty $\sigma_{r,i}$ corresponds to the fraction of the total uncertainty that is uncorrelated between two measurements that differ by one feature. The subscript i denotes an analysis feature difference, for instance one analysis using reconstruction Re1 and the other one using Re4. The first two terms of the off-diagonal terms correspond to the square

Table 2: Relative uncorrelated uncertainties $\sigma_{r,i}$ for a subset of ω_a analysis feature differences, for either A or T method category measurements. The relative precision of the reported numbers is estimated to be order 10% relative, and the amount of reported digits has been chosen for presentation convenience.

method category A			
Reco	Re1	Re2	Re3
Re2	0.03788		
Re3	0.03157	0.02020	
Re4	0.08334	0.07450	0.07955

method category T			
Reco	Re1	Re2	Re3
Re2	0.03632		
Re3	0.02951	0.02270	
Re4	0.18728	0.18047	0.18160

method category A		
Ratio	Ra1	Ra2
Ra2	0.07730	
Ra3	0.09289	0.07606

method category T		
Ratio	Ra1	Ra2
Ra2	0.06421	
Ra3	0.08393	0.07266

of the smaller of the two uncertainties, and when the third term is not present it corresponds to a critical correlation. With this covariance, the uncertainty on the residual between two measurements with $\sigma_B > \sigma_A$ is:

$$\sigma^2_{B-A} = \sigma^2_B - \sigma^2_A + \left(\sum_i \sigma^2_{r,i}\right) \cdot \left(\sigma^2_A + \sigma^2_B\right) \ . \tag{4}$$

The model assumes that the amount of uncorrelated uncertainty component related to the flavour of "Reco" method is independent of the uncorrelated uncertainty component related to the flavour of "Ratio" method. Table 2 reports measurements of a subset of the relative uncorrelated uncertainties $\sigma^2_{r,i}$ for ω_a measurements that differ by a single feature. The relative uncorrelated uncertainties have been obtained by measuring the spread of the residual of ω_a measurements with analyses with one feature difference on ∼200 bootstrap [5] samples, assembled by randomly extracting >200 subsamples of our datasets, with replacement. The measured spreads with real data bootstrap samples confirm previous studies for the "Ratio" feature differences, performed with simplified simulated toy Monte Carlo samples. They also confirm on real data that the measurements using the A and T method categories are consistent with being critically correlated. Finally, they are consistent with the model hypothesis that the amount of uncorrelated uncertainty for a feature difference is a constant fraction of the total uncertainty.

4. Consistency of the ω_a measurements

We use the modeled covariance to obtain pulls for all ω_a measurements using the same dataset, for all the datasets used for the muon $g-2$ measurement published in 2023. When dedicated studies for relative uncorrelated uncertainties are missing for a feature difference, we assume that the information for the A method category can be used for the Q method category measurements, and that when considering analyses with differing method categories one can use the geometric average of the measured relative uncorrelated uncertainties. The resulting correlation of the ω_a

Table 3: Statistical correlation of the ω_a measurements. Each row and column identify a measurement analysis group, followed by the method category prepended with "R" for measurements using the "Ratio" technique, in order to distinguish measurements made by the same group, with the same method category, and with or without the "Ratio" technique. More details on each measurement features are reported in Table 1. The amount of reported digits has been set for the convenience of the presentation and does not imply a corresponding precision of the reported numbers.

	Gr2-T	Gr3-T	Gr4-T	Gr5-T	Gr6-T	Gr1-A	Gr2-A	Gr3-A	Gr4-A	Gr5-A	Gr6-A	Gr1-RT	Gr3-RT	Gr4-RT	Gr1-RA	Gr3-RA	Gr7-Q	Gr7-RQ
Gr1-T	0.967	0.999	0.967	0.999	1.000	0.900	0.871	0.884	0.867	0.884	0.884	0.993	0.995	0.963	0.895	0.904	0.765	0.824
Gr2-T		0.967	1.000	0.965	0.967	0.891	0.900	0.875	0.896	0.874	0.875	0.961	0.963	0.996	0.887	0.895	0.756	0.815
Gr3-T			0.967	0.999	0.999	0.913	0.885	0.898	0.880	0.898	0.897	0.993	0.996	0.963	0.909	0.918	0.753	0.811
Gr4-T				0.965	0.967	0.897	0.906	0.881	0.902	0.880	0.880	0.961	0.963	0.996	0.892	0.901	0.751	0.809
Gr5-T					0.999	0.915	0.886	0.900	0.882	0.902	0.899	0.992	0.995	0.961	0.911	0.920	0.751	0.809
Gr6-T						0.915	0.887	0.899	0.882	0.899	0.899	0.993	0.995	0.963	0.911	0.919	0.752	0.810
Gr1-A							0.994	1.000	0.994	0.999	1.000	0.890	0.886	0.887	0.991	0.994	0.688	0.740
Gr2-A								0.994	1.000	0.993	0.994	0.862	0.857	0.896	0.986	0.988	0.681	0.732
Gr3-A									0.994	0.999	1.000	0.875	0.871	0.871	0.991	0.994	0.676	0.727
Gr4-A										0.993	0.994	0.858	0.853	0.892	0.986	0.988	0.678	0.729
Gr5-A											0.999	0.875	0.871	0.870	0.990	0.993	0.677	0.728
Gr6-A												0.875	0.870	0.871	0.991	0.994	0.676	0.727
Gr1-RT													0.994	0.962	0.902	0.907	0.758	0.825
Gr3-RT														0.967	0.895	0.901	0.767	0.837
Gr4-RT															0.895	0.901	0.750	0.819
Gr1-RA																0.994	0.682	0.743
Gr3-RA																	0.689	0.754
Gr7-Q																		0.994

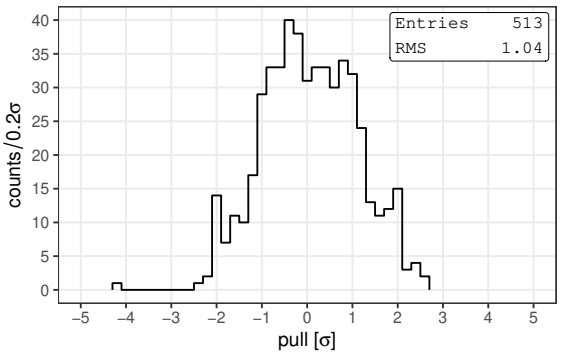

Figure 1: Residual pulls for ω_a measurements on the Run-2, Run-3a and Run-3b datasets.

measurements is reported in Table 3. In order to compare also measurements that only differ by additional minor analysis features other than the "Ratio" and "Reco" features, the residual uncertainty has been increased in quadrature by an additional estimated systematic uncertainty of 28 ppb. Figure 1 reports the observed pulls, which are consistent with being Normally distributed. There is one single outlier with $|\text{pull}| > 3$, correspondig to the pull between two "Ratio" features of Q method category measurements on one dataset. That is considered acceptable, since there have been no studies for this feature difference in Q method category measurements, and its corresponding relative uncorrelated uncertainty has been estimated to be the same as the one measured for the A measurements. Furthermore, as detailed in the next section, the Q (and T) method category measurements are not used to compute the per-dataset ω_a averages.

5. Averaging of the ω_a measurements

Although the estimated covariance for the ω_a measurements might be used to compute the optimal average of all ω_a measurements on each dataset, the resulting average of these highly correlated measurements would be very sensitive to the uncertainties on the measured correlations, as mentioned in Section 7.6.2 of Ref. [2]. The per-dataset ω_a averages are therefore obtained by evenly averaging one representative measurement performed by each of the 6 analysis groups that performed the most precise A-method-category measurements. This choice is expected to minimize the total systematic uncertainty of the measurement. The uncertainty on this average is computed by using a conservative covariance corresponding to 100% correlation between the 6 averaged measurements. Using the estimated covariance for the ω_a measurements we are able to estimate that the statistical uncertainty of the even average with the conservative covariance is only about 1.5% larger than the optimal statistical uncertainty.

6. Acknowledgments

This work was supported in part by the US DOE, Fermilab, the Italian Istituto Nazionale di Fisica Nucleare (INFN), and the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101006726.

References

[1] Muon g-2 collaboration, D. P. Aguillard et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm*, *Phys. Rev. Lett.* **131** (2023) 161802, [2308.06230].

[2] G. Cowan, *Statistical data analysis*. Oxford University Press, USA, 1998.

[3] Muon g-2 collaboration, G. W. Bennett et al., *Statistical equations and methods applied to the precision muon (g-2) experiment at BNL*, *Nucl. Instrum. Meth. A* **579** (2007) 1096–1116.

[4] Muon g-2 collaboration, D. P. Aguillard et al., *Detailed report on the measurement of the positive muon anomalous magnetic moment to 0.20 ppm*, *Phys. Rev. D* **110** (2024) 032009, [2402.15410].

[5] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, *Annals Statist.* **7** (1979) 1–26.