

Fast fits for lattice QCD correlatorsK. Hornbostel,¹ G. P. Lepage,^{2,*} C. T. H. Davies,³ R. J. Dowdall,³ H. Na,⁴ and J. Shigemitsu⁴(HPQCD collaboration)[†]¹*Southern Methodist University, Dallas, Texas 75275, USA*²*Laboratory of Elementary-Particle Physics, Cornell University, Ithaca, New York 14853, USA*³*SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK*⁴*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*

(Received 8 November 2011; published 8 February 2012)

We illustrate a technique for fitting lattice QCD correlators to sums of exponentials that is significantly faster than traditional fitting methods—10–40 times faster for the realistic examples we present. Our examples are drawn from a recent analysis of the Υ spectrum, and another recent analysis of the $D \rightarrow \pi$ semileptonic form factor. For single correlators, we show how to simplify traditional effective-mass analyses.

DOI: 10.1103/PhysRevD.85.031504

PACS numbers: 11.15.Ha, 12.38.Gc

Most physics results in lattice QCD come from fits of lattice correlators to sums of exponentials. For example, we study a particular hadron by computing Monte Carlo simulation estimates $G_{ab}^{\text{MC}}(t)$ of hadronic correlators,

$$\sum_{\mathbf{x}} \langle 0 | \Gamma_b(\mathbf{x}, t) \Gamma_a(0, 0) | 0 \rangle, \quad (1)$$

with different sources Γ_a and sinks Γ_b that create and destroy the hadron. The sum over all spatial sites \mathbf{x} restricts the hadrons to states with zero total three-momentum. Such a correlator can be decomposed into contributions from energy eigenstates $|E_j\rangle$ in QCD [1]:

$$G_{ab}(t; N) = \sum_{j=1}^N a_j b_j \exp(-E_j t) \quad (2)$$

where E_j is the energy, with $E_j \geq E_{j-1}$, and the amplitudes are matrix elements, with

$$a_j^* = \langle 0 | \Gamma_a(0, 0) | E_j \rangle, \quad b_j = \langle 0 | \Gamma_b(0, 0) | E_j \rangle. \quad (3)$$

The physics is in the energies and the matrix elements, and these are determined by fitting formula (2) to the Monte Carlo data $G_{ab}^{\text{MC}}(t)$ for a variety of sources and sinks.

In principle, the number of terms N in Eq. (2) is infinite, but, in practice, we need only retain a finite number of terms because the exponentials suppress high-energy states. The number needed depends upon the precision of the simulation data G_{ab}^{MC} , but it is not uncommon to require $N = 10$ or more terms for good fits to accurate data. The fitting process becomes both cumbersome and time consuming if many correlators must be fit simultaneously while using such large N s. In this paper we introduce a

method that can dramatically simplify and accelerate such fits.

The key to this new approach lies in how priors are introduced. Two types of input data are required for these fits. The first is simulation data for the correlators, consisting of Monte Carlo averages \bar{G} for each a, b and t , and a covariance matrix σ^2 that specifies both the statistical uncertainties in each average and the correlations between them:

$$G_{ab}^{\text{MC}}(t) \leftrightarrow \left\{ \bar{G}_{ab}(t), \sigma_{ab,a'b'}^2(t, t') \right\} \quad (4)$$

This data contributes

$$\begin{aligned} \chi_{\text{MC}}^2(a_j, b_j, E_j) = & \sum_{t,a,b,t',a',b'} (G_{ab}(t; N) - \bar{G}_{ab}(t)) \sigma_{ab,a'b'}^{-2}(t, t') \\ & \times (G_{a'b'}(t'; N) - \bar{G}_{a'b'}(t')) \end{aligned} \quad (5)$$

to the χ^2 function that is minimized by varying parameters a_j , b_j , and E_j in a conventional fit.

The second type of input data consists of Bayesian priors for each fit parameter. Complicated multi-correlator, multi-parameter fits are impossible without *a priori* estimates for each fit parameter [2,3]:

$$a_j^{\text{pr}} \equiv \bar{a}_j \pm \sigma_{a_j}, \quad b_j^{\text{pr}} \equiv \bar{b}_j \pm \sigma_{b_j}, \quad E_j^{\text{pr}} \equiv \bar{E}_j \pm \sigma_{E_j}. \quad (6)$$

This information is included in a conventional fit by adding extra terms to $\chi^2(a_j, b_j, E_j)$: $\chi^2 = \chi_{\text{MC}}^2 + \chi_{\text{pr}}^2$ where

$$\chi_{\text{pr}}^2(a_j, b_j, E_j) = \sum_{j=1}^N \left[\frac{(a_j - \bar{a}_j)^2}{\sigma_{a_j}^2} + \frac{(b_j - \bar{b}_j)^2}{\sigma_{b_j}^2} + \frac{(E_j - \bar{E}_j)^2}{\sigma_{E_j}^2} \right]. \quad (7)$$

The priors can also be combined to give *a priori* estimates for the correlators,

*g.p.lepage@cornell.edu

†http://www.physics.gla.ac.uk/HPQCD

$$G_{ab}^{\text{pr}}(t; N) \equiv \sum_{j=1}^N a_j^{\text{pr}} b_j^{\text{pr}} \exp(-E_j^{\text{pr}} t), \quad (8)$$

where the means and covariance matrix for $G_{ab}^{\text{pr}}(t)$ are computed, using standard error propagation, from the means and covariance matrix of the priors (Eq. (6)).

The cost of a traditional analysis goes up rapidly with the number of parameters needed to obtain a good fit. Our new approach takes advantage of the fact that we are rarely interested in the values of parameters from large- j terms in fit function (2), even when these terms are needed for a good fit. Rather than including them in the fit, we instead incorporate the large- j terms into the Monte Carlo data *before* fitting. This reduces the number of fit parameters, leading to much faster fits.

We incorporate large- j terms into the Monte Carlo data by using the priors to generate *a priori* estimates for these terms (including the uncertainties from the priors), which we then subtract from the Monte Carlo data. This effectively removes the large- j terms from the data. The modified data is then fit with a simpler formula that includes only small- j terms.

More explicitly, we remove terms having $n < j \leq N$ by replacing $G_{ab}^{\text{MC}}(t)$ with (first definition)

$$\tilde{G}_{ab}^{\text{MC}}(t; n) \equiv G_{ab}^{\text{MC}}(t) - \Delta G_{ab}^{\text{pr}}(t; n), \quad (9)$$

where

$$\Delta G_{ab}^{\text{pr}}(t; n) \equiv G_{ab}^{\text{pr}}(t; N) - G_{ab}^{\text{pr}}(t; n) = \sum_{j=n+1}^N a_j^{\text{pr}} b_j^{\text{pr}} \exp(-E_j^{\text{pr}} t) \quad (10)$$

is the $j > n$ part of the fit function. Having removed the $j > n$ terms, we fit $\tilde{G}_{ab}^{\text{MC}}(t; n)$ with the simpler fit function, $G_{ab}(t; n)$, rather than $G_{ab}(t; N)$.

Here we assume that N is sufficiently large that $\Delta G_{ab}^{\text{pr}}(t; n)$ and therefore $\tilde{G}_{ab}^{\text{MC}}(t; n)$ are independent of N to within their statistical errors. The covariance matrix for $\tilde{G}_{ab}^{\text{MC}}(t; n)$ is obtained by adding the covariance matrices of $G_{ab}^{\text{MC}}(t)$ and $\Delta G_{ab}^{\text{pr}}(t; n)$ (that is, adding the errors in quadrature) [4].

Removing high- j terms from both the fit function and the fit data replaces the original fitting problem—fit an N -term function $G_{ab}(t; N)$ to $G_{ab}^{\text{MC}}(t)$ —by a simpler problem that can have far fewer fit parameters: fit an n -term function $G_{ab}(t; n)$ to $\tilde{G}_{ab}^{\text{MC}}(t; n)$, where $n < N$. Remarkably, as we showed in [5], these two problems are equivalent for high statistics data even when n is quite small: that is, fit results (means and standard deviations) for the low- j parameters are the same in both cases. In the second case, the $j > n$ terms have been “marginalized,” or, in effect, integrated out of the Bayesian probability distribution, but in a way that does not affect the analysis of the $j \leq n$ terms. When $n \ll N$, the fit parameters that remain are many

fewer than what would be required in a standard fit, and fitting is much faster.

The new analysis, based on Eq. (9), is only approximately equivalent to the standard analysis—that is, only insofar as the various probability distributions involved can be approximated by Gaussian distributions. As a result very small values of n may not work well. The rapid exponential falloff of the propagators exacerbates this problem here, which suggests that we replace definition (9) by

$$\log \tilde{G}_{ab}^{\text{MC}}(t; n) \equiv \log G_{ab}^{\text{MC}}(t) - \Delta \log G_{ab}^{\text{pr}}(t; n), \quad (11)$$

where

$$\Delta \log G_{ab}^{\text{pr}}(t; n) \equiv \log G_{ab}^{\text{pr}}(t; N) - \log G_{ab}^{\text{pr}}(t; n). \quad (12)$$

Rearranging and exponentiating, this variation gives modified propagators that are defined by (second definition)

$$\tilde{G}_{ab}^{\text{MC}}(t; n) \equiv G_{ab}^{\text{MC}}(t) \frac{G_{ab}^{\text{pr}}(t; n)}{G_{ab}^{\text{pr}}(t; N)}, \quad (13)$$

We will use this second implementation of the marginalization procedure throughout the rest of this paper, since we find that it gives good results for values of n that are two or three times smaller than those from the first implementation. Again, terms with $j > n$ have been removed, and therefore the modified correlator data is fit with the simpler fit function, $G_{ab}(t; n)$.

We now illustrate our new method by applying it to QCD simulation data from two recent analyses. For each analysis, we fit a function, like $G_{ab}(t; n)$, with n terms both to untouched simulation data $G_{ab}^{\text{MC}}(t)$, and to modified simulation data $\tilde{G}_{ab}^{\text{MC}}(t; n)$, from which $j > n$ terms have been removed using Eq. (13). We vary n , doing sequential fits with $n = 1, 2, 3, \dots$, where the best-fit parameter values from one fit are used as starting values for the next fit. Sequential fitting with increasing n is a standard approach to complicated multi-parameter correlator fits; n is increased until the fit's χ^2 stops changing, at which point enough terms have been included to reflect accurately the uncertainties introduced by large- j terms. Here we examine the best-fit parameters for each n to investigate the rate at which correct results emerge from this process. This allows a detailed comparison of our two fitting strategies.

The first data set is a collection of 25 correlators for the $Y(1S)$ meson and its radial excitations ($Y(2S)$, $Y(3S)$, *etc.*) [6]. These correlators were made using five different operators for both sources and sinks. They were fit to formula (2) with priors (in lattice units):

$$\log(E_1) = \log(0.3 \pm 0.1) = -1.2 \pm 0.3$$

$$\log(E_{j+1} - E_j) = \log(0.25 \pm 0.125) = -1.4 \pm 0.5 \quad (14)$$

$$a_j = 0.1 \pm 1.0$$

except for a local source for which the priors were

FAST FITS FOR LATTICE QCD CORRELATORS

$\log(a_j) = \log(0.1 \pm 0.2) = -2.3 \pm 2$ (local source). These are broad priors—more than 100 times broader than the final errors for the quantities we examine below. We set $N = 20$ when defining $\tilde{G}_{ab}^{MC}(t; n)$ (Eq. (13)); this is roughly twice the size it needs to be, but it costs little to make N large. In general N should be chosen so that terms with $j > N$ are negligible compared with statistical errors.

In Fig. 1 we plot the χ^2 per degree of freedom for each method versus the time required to get to that value [7]. As expected, the new algorithm reaches a reasonable χ^2 with just a few terms ($n = 2-3$), in 20–30 seconds; the traditional algorithm requires $n = 10-11$ to obtain a good χ^2 , and 600–700 seconds. Similar differences are evident if we look at physical quantities extracted from the simulations. In Fig. 2 we show results for the $2S - 1S$ mass splitting (in lattice units), for the $3S - 1S$ mass splitting divided by the $2S - 1S$ splitting, and for the $1S$ and $2S$ mesons' (non-relativistic) wave functions at the origin, which come from fit parameters a_j for a local source. In every case the two algorithms agree on the final result, but the new algorithm converges to correct results 10–40 times faster.

Our second example is from a recent analysis of the $D \rightarrow \pi$ semileptonic form factor [8]. To extract the form factor at four different momenta, this analysis uses a simultaneous fit of 13 two-point and three-point correlators: a) a D -meson correlator with a pseudoscalar local source and sink; b) four π -meson correlators, one for each pion momentum of interest, again with local pseudoscalar sources and sinks; and c) two three-point correlators $D \rightarrow J_{\text{scalar}} \rightarrow \pi$ for each of the four pion momenta. The fit functions are more complicated for this case. For example, the D -meson correlator is fit by a function:

$$G_D(t; n) = \sum_{j=1}^n a_j f(E_j, t) - (-1)^j a_j^o f(E_j^o, t) \quad (15)$$

where $f(E_j, t) \equiv \exp(-E_j t) + \exp(-E_j(T - t))$ is periodic with period $T = 64$, and the second (oscillating)

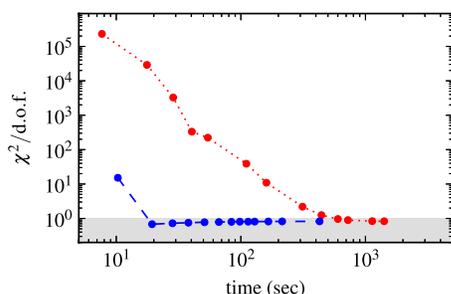


FIG. 1 (color online). Fit χ^2 per degree of freedom for sequential fits of 25 Y correlators with $n = 1, 2, 3 \dots$ terms in fit function (2). Results are plotted versus the cumulative time required for fitting, and are for fits of: a) the unmodified simulation data $G_{ab}^{MC}(t)$ (red circles and dotted line); and b) the modified simulation data $\tilde{G}_{ab}^{MC}(t; n)$ (Eq. (13)) (blue circles and dashed line). The region of good fits is indicated by the gray band.

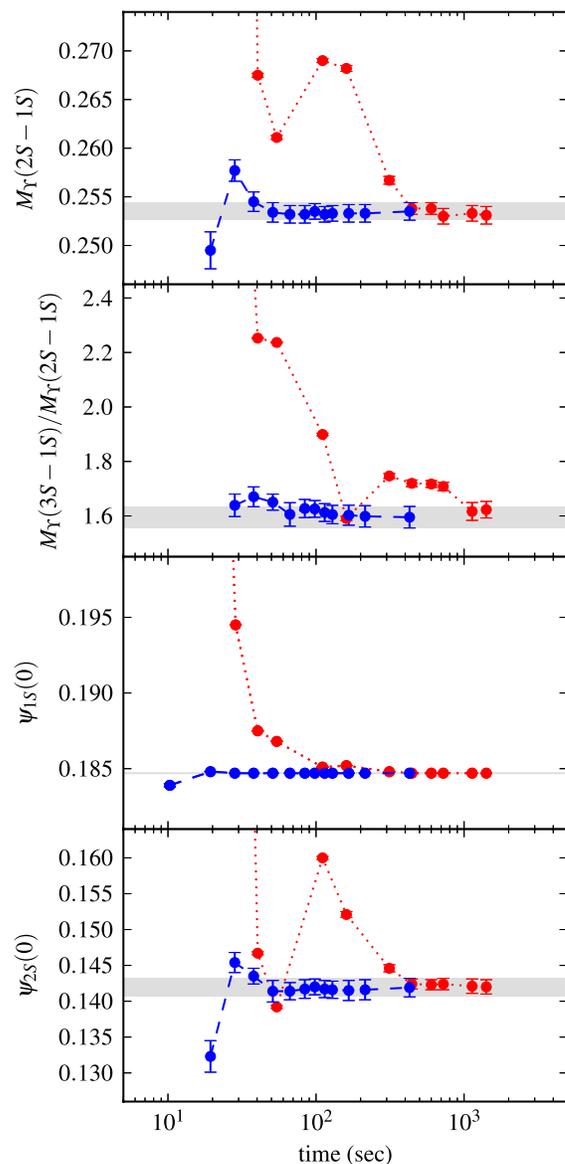
PHYSICAL REVIEW D **85**, 031504(R) (2012)

FIG. 2 (color online). Best-fit results from sequential fits of 25 Y correlators with $n = 1, 2, 3 \dots$ terms in fit function (2). Results are plotted versus the cumulative time required for fitting, and are for fits of: a) the unmodified simulation data $G_{ab}^{MC}(t)$ (red circles and dotted line); and b) the modified simulation data $\tilde{G}_{ab}^{MC}(t; n)$ (Eq. (13)) (blue circles and dashed line). Results are given for mass splittings between different vector S -states, and for the wave functions at the origin for the lowest two states. All results are in lattice units. The gray bands show the best-fit result from the modified data after convergence.

term is due to opposite-parity states in the correlator (a feature of staggered-quark formalisms like that used in this analysis). The details for the other correlators, and the priors are given in [8].

Despite the complexity of dealing with both two-point and three-point correlators, this is a simpler fit than the Y case; but even here we find that marginalizing most of the fit function makes the analysis about 30 times faster. We

show results in Fig. 3 for the D -meson's mass m_D and leptonic decay constant f_D , as well as for the $D \rightarrow \pi$ scalar form factor $f_0(0, 0, 0)$ at zero recoil momentum. All results are in lattice units. Again the two approaches agree on the results but the new approach has correct results even with only a single term ($n = 1$) in the fit functions. For these fits we set $N = 10$ when computing the modified data $\tilde{G}_{ab}^{\text{MC}}(t; n)$ (Eq. (13)), which is twice as large as it needs to be.

Some insight into how marginalization works can be gained by focusing just on the D correlator from this analysis and fitting the modified data,

$$\tilde{G}_D^{\text{MC}}(t) \equiv G_D^{\text{MC}}(t) \frac{a_1^{\text{pr}} f(E_1^{\text{pr}}, t)}{G_D^{\text{pr}}(t; N)}, \quad (16)$$

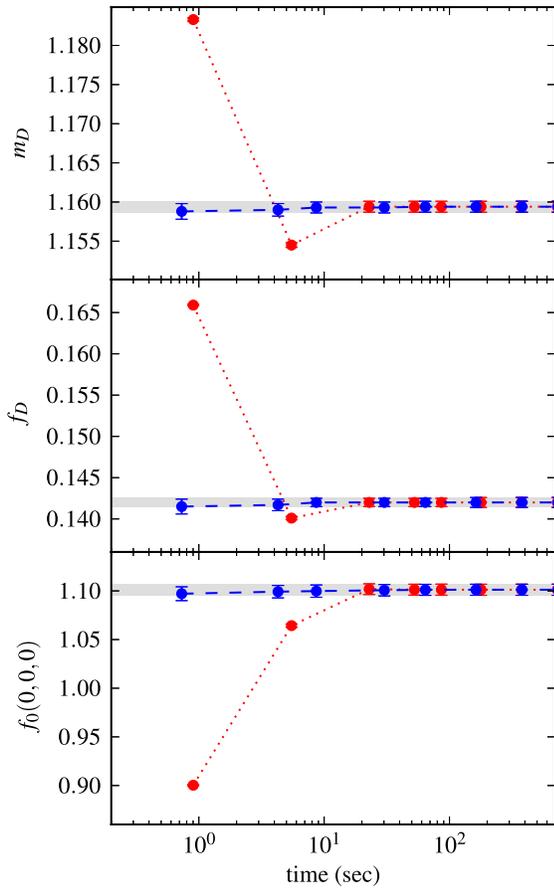


FIG. 3 (color online). Best-fit results from sequential fits of 13 two-point and three-point correlators for D and π mesons with $n = 1, 2, 3 \dots$ terms in fit function (2). Results are plotted versus the cumulative time required for fitting, and are for fits of: a) the unmodified simulation data (red circles and dotted line); and b) the modified simulation data (Eq. (13)) (blue circles and dashed line). Results are given for the D -meson mass m_D and decay constant f_D , and for the $D \rightarrow \pi$ scalar form factor at zero recoil momentum $f_0(0, 0, 0)$. All results are in lattice units. The gray bands show the best-fit result from the modified data after convergence.

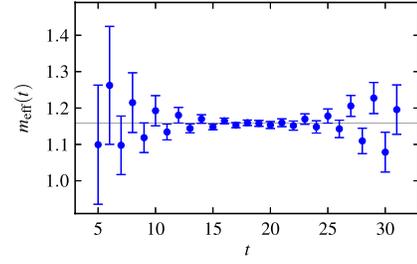


FIG. 4 (color online). The D -meson's effective mass $m_{\text{eff}}(t)$ versus t computed from modified simulation data $\tilde{G}_D^{\text{MC}}(t)$ from which every state other than the ground state has been removed (using priors). The (very thin) gray band shows the weighted average of all $m_{\text{eff}}(t)$ s, taking account of correlations. The thickness of the band indicates the uncertainty of the average. Note that the largest ts shown here correspond to the middle t range. The error bars grow there because $m_{\text{eff}}(t)$ becomes very sensitive to statistical errors in this region (since periodic boundary conditions imply that the derivative of the correlator's non-oscillating part vanishes at the midpoint).

with only the non-oscillating part of the first term in Eq. (15)—that is, with $a_1 f(E_1, t)$. This situation is sufficiently simple that fitting is not required. The D mass, for example, can be obtained by averaging the “effective mass,”

$$m_{\text{eff}}(t) \equiv \text{arccosh} \left(\frac{\tilde{G}_D^{\text{MC}}(t+1) + \tilde{G}_D^{\text{MC}}(t-1)}{2\tilde{G}_D^{\text{MC}}(t)} \right), \quad (17)$$

over all t , taking account of correlations between different ts . The effective mass is plotted as a function of t in Fig. 4. It is compared with the weighted average of all 27 $m_{\text{eff}}(t)$ s (gray band), which at $m_{\text{eff}}^{\text{avg}} = 1.1584(11)$ agrees well with the best result, $1.1593(7)$, from full multi-term fits (top panel in Fig. 3).

The first excited state in the D correlator is the opposite-parity contribution, which accounts for the oscillation in $m_{\text{eff}}(t)$. Strong statistical correlations between different points result in an average m_{eff} whose error is more than 7 times smaller than the best error from an individual $m_{\text{eff}}(t)$. The errors in $m_{\text{eff}}(t)$ when $t \leq 16$ come almost entirely from marginalized terms absorbed into the fit data using Eq. (16); the original Monte Carlo simulation errors are negligible there.

In the absence of marginalization, contributions from excited states would limit a traditional effective mass analysis of this data to values with $t > 16$. With marginalization, all ts are used, except for a small number at very small t where the fit function is invalid (because of temporal non-locality in the lattice quark action). Using 28 ts is possible because we have removed the excited states through Eq. (16). As a result different $m_{\text{eff}}(t)$ s agree with each other to within their errors: fitting all 27 values in Fig. 4 to a constant gives an excellent fit, with a χ^2 per degree of freedom of 0.6. (The result of the fit is, by definition, the same as the weighted average reported above.)

Our new implementation of effective-mass analyses is simpler and less ambiguous than traditional analyses because we are not limited to large ts . More importantly our implementation also allows us to quantify the contribution to the uncertainty in the final $m_{\text{eff}}^{\text{avg}}$ due to the excited states: here the priors for non-oscillating terms in Eq. (15) contribute $0.44\sigma_m$, those from oscillating terms contribute $0.07\sigma_m$, and the uncertainties in the Monte Carlo data contribute $0.89\sigma_m$, where σ_m is the standard deviation of $m_{\text{eff}}^{\text{avg}}$. Such information is essential for assessing the reliability of the final result, as well as for planning improvements to the analysis.

In this paper we have shown how to accelerate multi-exponential fits to multiple hadronic correlators by removing contributions due to excited states from both the fit function and the simulation data, before fitting.

This technique for marginalizing large parts of the fit function greatly reduces the number of fit parameters needed in the realistic examples presented here, and makes fitting 10–40 times faster. Marginalization also simplifies effective-mass analyses, and generalizes easily to analogous multi-state (generalized eigenvalue) methods.

This work was supported by the DOE (DE-FG02-04ER41299, DE-FG02-91ER40690), the NSF (PHY-0757868), and the STFC. We used the Darwin Supercomputer of the Cambridge High Performance Computing Service as part of the DiRAC facility jointly funded by STFC, BIS and the Universities of Cambridge by the Office of Science of the DOE and at the Ohio Supercomputer Center.

-
- [1] Lattice QCD simulations use Euclidean time and so $-it$ is replaced by $-t$ in the exponentials. Also simulations are for finite volumes in space, and therefore all states, including multi-hadron states, have discrete energy eigenvalues.
- [2] See G. P. Lepage, B. Clark, C. T. H. Davies, K. Hornbostel, P. B. Mackenzie, C. Morningstar, and H. Trotter, *Nucl. Phys. B, Proc. Suppl.* **106-107**, 12 (2002). The formula for χ_{pr}^2 generalizes trivially if there are correlations between the priors for different parameters. Since the priors are new input data, leading to new terms in the χ^2 function, the number of degrees of freedom in the fit is the number of pieces of original Monte Carlo data plus the number of priors minus the number of fit parameters. Consequently the number of degrees of freedom always equals the number of pieces of Monte Carlo data since there is a prior for each fit parameter. This is true however many parameters are included, even when the number parameters exceeds the number of data points in the original Monte Carlo data. In practice one adds terms until the fit results (means and standard deviations, $\chi^2 \dots$) converge. Adding further terms has no effect, but reassures us that systematic errors due to truncation of the fit function are negligible.
- [3] In practice useful priors are easily found. As examples, see the extended analyses in: C. T. H. Davies *et al.* (HPQCD Collaboration), *Phys. Rev. D* **78**, 114507 (2008); C. McNeile *et al.* (HPQCD Collaboration), *Phys. Rev. D* **82**, 034512 (2010). Incorrect priors are immediately evident since the χ^2 per degree of freedom (Fig. 1, for example) does not converge to a reasonable value (of order 1 or less) as more fit parameters are included. This is apparent much more quickly using the new method presented here since convergence occurs with many fewer parameters.
- [4] Again the covariance matrix for $\Delta G_{ab}(t; n)$ is computed using standard error propagation—for example, $f(\bar{x} \pm \sigma_x) = \bar{f} \pm \sigma_f$ with $\bar{f} \approx f(\bar{x})$ and $\sigma_f^2 \approx f'(\bar{x})^2 \sigma_x^2$. We have compared this linearized analysis with Monte Carlo evaluations of ΔG (from normal distributions for the priors). We find the Monte Carlo results to be both much more expensive and also less robust for correlators that decay exponentially quickly. Note also that it is essential to retain the off-diagonal elements (correlations) in the covariance matrix for $\Delta G_{ab}(t; n)$; correlations arise because, for example, the prior data used for a parameter is the same for all t values.
- [5] For a proof, see the appendix of C. McNeile, C. T. H. Davies, E. Follana, K. Hornbostel, and G. P. Lepage, *Phys. Rev. D* **82**, 034512 (2010).
- [6] The simulations used 0.09 fm lattices with $n_f = 4$ sea quarks (HISQ discretization), and NRQCD dynamics for the b quark. The gluon configurations were provided by the MILC collaboration. For further details see: R. J. Dowdall *et al.*, [arXiv:1110.6887](https://arxiv.org/abs/1110.6887).
- [7] The absolute computer times quoted here are obviously of little relevance since they depend upon specific details of hardware and software. What is relevant is the comparison between methods.
- [8] The simulations used 0.12 fm lattices with $n_f = 3$ light sea quarks (ASQTAD discretization), and HISQ relativistic dynamics for valence quarks. The gluon configurations were provided by the MILC collaboration. For further details see (set C2): H. Na, C. T. H. Davies, E. Follana, J. Koponen, G. P. Lepage, and J. Shigemitsu, *Phys. Rev. D* **84**, 114505 (2011).