Supplement 1 to “Criticality of adaptive control dynamics”:

Scaling and significance

Felix Patzelt and Klaus Pawelzik

Institute for Theoretical Physics, University of Bremen, Germany

(Dated: December 2, 2011)

Abstract

We discuss quantifying the goodness of fits for power-law tails. A method to calculate p-values by comparison to appropriate surrogate data is presented. Using this method, it is not possible to reject the hypothesis, that the tails of error distributions from the continuous model and experimental time series follow power-laws. It is demonstrated, that spectral densities allow to reject alternative models or parameter choices. Finally, scaling properties of the variance and mean displacements of control errors over different time scales are discussed.
GOODNESS OF FITS FOR THE PROBABILITY DENSITY

As ubiquitous as power-laws are in nature, as notoriously difficult they are to estimate and validate. This difficulty lies in the very nature of the analysis of distribution tails, since statistics have to be made for intrinsically rare events. Strong deviations from the ideal linear appearance of a power-law in log-log coordinates therefore can occur just by chance. Hence, large numbers of samples are needed to reliably characterize a power law. The power-law has also to be observed over more than an order of magnitude on both axes to distinguish it from other fat-tailed distributions. Even then, it may be possible, that the underlying distribution is, e.g. a combination of five different exponentials that together give the appearance of scaling. In this case, however, saying that the distribution is well described by a power-law still gives a good characterization of the effective observations that is practically impossible to distinguish from the true distribution, but uses less parameters.

The probability, that the fluctuations in a given data set occur by chance for a real power-law can be assessed using significance tests like the Kolmogorov-Smirnov (KS) test. However, there are also systematic sources for deviations. Experimental data often is only approximated by a certain analytical distribution of independent random variables, but does not truly follow this idealization. In very large data sets, even the smallest differences may become significant. A common source for such deviations are correlations, including higher order ones. Then, the measurements are not perfectly independent anymore and deviations are larger than analytically predicted for absolutely independent random variables. Another cause for deviations from an analytical power law in finite data sets is, that often only the tail of a distribution follows a power-law in the limit of very large events. Then, when determining a cutoff above which to fit a non-trivial tradeoff has to be made between systematic deviations due to imperfect convergence of smaller events and random fluctuations for the largest events. Last not least it is possible, that a distribution has no hard cutoff, but a slow gradual convergence towards a power-law.

To assess how well of an idealization a power-law is for the model and experimental error distributions, we proceed in three steps. First, we show, how deviations from the power-law in the model time series depend on correlations. Second, we will introduce a method to compare the goodness of fits between a given distribution and surrogate data. This is done by comparing how well model time series and power-law distributed independent random
variates are fitted by an analytical power-law. Third, we perform the same comparison between model and experimental time series. The reasoning behind this proceeding will become clear in the following.

Fig. 1 (a) shows the distribution of balancing errors $F_c(|y|)$ made by the continuous control model, and a fit. As explained in the main paper, the distribution is in double logarithmic coordinates approximately linear for large events, but takes a sudden break at the largest one. The reason for this effect is, that the largest few data points all belong to a single peak in the time series. To reduce the correlations between the analyzed events, distributions of subsets of the same time series are shown where only one event every second (Fig. 1 (b)) and every 10 seconds (Fig. 1 (c)) were used [6]. Finally, Fig. 1 (d) shows the distribution for independent random variates that follow a power-law above a threshold. A method to quantitatively evaluate the goodness of the respective fits is the KS-statistic. This nonparametric test calculates the maximum distance $D$ between two distributions. For independent random variates, the p-values can then be calculated independently of the actual shape of the distributions that are tested against each other. Here, we call these common p-values $p_{KS}$ values. They are stated for each respective subplot in Fig. 1. The probability, that the deviations of model distributions from an analytical power-law are explained by chance alone increases as the correlations decrease. When only one event every 10s is taken into account, the $p_{KS}$-value for this particular time series even reaches 0.9. That is, even though the largest approximately 5 data points appear to deviate from the power law, even larger deviations are to be expected by chance alone according to the KS-statistics. On average, of course, one expects p-values values of 0.5 for two identical distributions. Commonly, p-values below 0.05 lead to the rejection of the hypotheses, that the observed deviations between two distributions can be assessed to chance alone.

The previous method has several disadvantages. First, massive subsampling is required. Applied to real data sets containing at most $10^6$ events, only $10^3$ events are left over. Since only few of these events belong to the distribution tail, it spans less than one order of magnitude of control errors. Second, reducing the data set reduces the discriminatory power of the KS-Test. Third, to use as many data points as possible in smaller data sets and avoid fitting random kinks, it is useful to calculate fits using a range of cutoffs and then use only the one that minimizes a loss function. We exclusively used the cutoff that minimizes the KS-D-value. From our experience, various criteria generally tend to underestimate the cutoff
if the tested range includes too small events. This additional complication may decrease the
goodness of some fits. Therefore, we determine p-values by numerically calculating the
probability, that the KS statistics of some surrogate data shows bigger deviations from
their respective fits than the tested distribution (a variant of this method without regarding
correlations is discussed in [1]). This p-value is called $p >$ in the following and in Fig. 1.
Again, we expect this value to be 0.5 on average for data sets drawn from exactly the same
distribution. In principle, this method requires no subsampling. Unfortunately, it is not
straight forward to obtain correlated power-law distributed random variates. Therefore, we
first compare very large subsampled model time series to independent random variates. We
then use model time series as surrogate data to test the experimental ones with and without
subsampling. [7]

Fig. 2 (a) Shows the average $p >$ for model time series with different amounts of sub-
sampling where the total number of events was held constant at three different values cor-
responding to the three curves. That is, correlations are reduced, without reducing the
number of analyzed events. Therefore, the observed convergence towards 0.5 for low corre-
lations is not caused by a loss of power of the test. Similar results are found for different
parameter sets including controller memory time constants $τ_m$ that minimize mean squared
control errors where pdf tail exponents $δ$ are close to five (not shown). As the number of
events is increased from the red curve to the blue one, the test becomes more discriminative
as expected (see discussion at the end of this section).

Fig. 2 (b) shows the realistic case, where the length of the analyzed time series’ is $10^6$
without subsampling and is then reduced as would be the case for experimental data. The
p-value again converges towards 0.5.

Fig. 3 shows the p-values for the experimental time series using the above method with
model time series as surrogate data. Results scatter around $p > = 0.5$ as would be expected
for identical processes. Almost all p-values for full and subsampled time series are bigger than
the significance level $p > = 0.05$ for either all combined trials of each subjects or the combined
trials for each subject for each day analyzed separately. Only for Subject 5 both, combined
time series of all trials before and after subsapling failed the significance test. Here, closer
investigation revealed, that the overall scaling of the time series on different days differed
more than for the other subjects, leading to much lower p-values for the combined series
than for each day.
Fig. 4 shows the distributions and the fits also used in Fig. 3 (a) for the combined time series of all trials for each subject. Even distributions with low p-values are still clearly broader than Gaussian distributions. It can further be seen, that KS-Statistics are by construction more sensitive to deviations close to the cutoff than further down the in the tail of the distribution. For subject 3 (initial data set 3a), the noticeable kink in the distribution vanishes if the very first trial is discarded (3a*). A second data set from this subject was recorded a month later (3b), which also does not show a pronounced kink. However, similar distributions as in 3a sometimes occur by chance for simulations of the model. Distributions for single days are qualitatively similar to the ones shown in Fig. 4, but exponents vary over a slightly wider range. These distributions are not shown due to their large number.

Summing up, we could not reject the hypothesis, that model error distributions follow a power-law in the limit for large errors if correlations are taken into account. Furthermore, we could not reject the hypothesis, that experimental error distributions typically are described by a power law just as good as those from the model.

**Further discussion**

Here, we discuss two possible misconceptions of the above results that might occur. First, the p-values in Fig. 2 (a) get smaller from the red curve to the blue one, i.e. as the number of samples is increased. Naively, one might expect a constant correlation length to become insignificant for sufficiently large time series, leading to the opposite effect. However, the KS-Test assumes, that the maximum distance $D$ between the empirical cumulative distribution function of $n$ iid samples of a random variable and the true underlying distribution scales as $D \propto 1/\sqrt{n}$. This scaling is violated for correlated samples, for which convergence is slower. Therefore, while $D$ still converges towards zero, calculated $p_{KS}$-values also converge towards zero instead of keeping an average of 0.5 as is the case for independent samples. This effect is illustrated in Fig. 5 (a) which shows average $p_{KS}$-values for a comparison of the tails of two model time series with each other for different amounts of sub-sampling. As in Fig. 2 (a), correlations lead to an underestimation of the amount of deviations expected for a given sample size and therefore low $p$-values. Convergence towards 0.5 is even slower. Most notably, for a fixed amount of subsampling p-values decrease when the number of samples is increased. This demonstrates, that the KS-test indeed becomes more sensitive
to correlations for larger sample sizes. We also found similar effects for correlated Gaussian random variates (not shown), but could not find an example where correlations become insignificant for large sample sets.

The second potential misconception is, that steep power-laws with large exponents $\delta$ of five and more are impossible to observe. While we performed the rigorous analysis above for typical parameters of the model, Figure 2 (b) in the main paper shows exponents beyond the experimentally observed range. It is indeed true, that to clearly identify such a scaling behaviour an enormous number of samples would be necessary that is unlikely to be available in an experimental data-set. However, Fig. 5 (b) shows a simulation of the continuous control model for parameters just outside of the realistic range that does indeed show scaling with a very steep slope over one order of magnitude of control errors. Because of the computational resources required to calculate even just this single distribution, we did not perform a rigorous statistical analysis for this set of parameters. The goal of this figure is merely to show, that the scaling behavior does not simply stop above the range that is experimentally observed or that minimizes mean squared balancing errors for realistic reaction times. Even if distributions with a power-law tail with exponents up to ten would hypothetically occur in experimental data sets, while not showing scaling over a significant regime for realistic data set sizes, they would still be identifiable by our methods as broader than a Gaussian and not as broad as the distributions we observed.
FIG. 1: (a): Complementary cumulative distribution of balancing errors $|y|$ made by the continuous control model described in the main paper with controller reaction time $t_r = 170\,\text{ms}$, gain $\gamma = 1.07$ and memory time constant $\tau_m = 140\,\text{ms}$. The system time constant was $\tau = 250\,\text{ms}$ and the noise level $\sigma = 5\,\text{s}^{-0.5}$. Time discretization was $h = 11.8\,\text{ms}$. Short diagonal line: a power-law in the pdf $p(y) = |y|^{-\delta}$ corresponds to $F_c(y) = |y|^{-\delta+1}$. Figure generation like in the main paper. Horizontal lines indicate the cutoff used for the Hill-estimator. The cutoff was the one of 100 logarithmically spaced ones with the best KS-statistics (e.g. minimal $D$). (b), (c): Distributions of subsets of the same time-series as in (a), but with only one value of $|y|$ every $\Delta t = 1$ and $\Delta t = 10$ seconds used respectively. (d) Independent random variates that are distributed according to a Gaussian below a threshold $x_{\text{th}} = 2.5$ and to a power-law with $\delta = 3.5$ above. The exact analytical shape of the distribution was obtained by requiring the pdf and its first derivative to be continuous at $x_c$. The variates were then obtained from uniform ones using inverse transform sampling. All subfigures were created like Fig. 1 (a) in the main paper. $p_{KS}$-values refer to tabulated $p$-values for the Kolmogorov-Smirnov statistics. For the $p_{>}$-values, 1000 time series with random variates like in (d) were generated and a power-law was fitted to each one. The $p_{>}$-values refer to the fraction of times that the KS-statistics of a fit was worse than for the one distribution depicted in the respective subfigure. As the correlations in the model time series decrease, both goodness of fit tests converge towards those for the i.i.d. variates.
FIG. 2: (color online) Probability, that the KS-statistics for power-law fits for independent random variates that really follow a power-law above a threshold is worse than for control errors from the continuous control model with different amounts of decorrelation by sub-sampling. Parameters were chosen as in Fig. 1. For each data point, $10^3$ model time-series were compared to $10^3$ different sets of random variates each. Dashed line: A $p_\geq$-value of 0.5 that would be expected, if model control errors and surrogate random variates were truly following the same distribution. Dotted line: $p_\geq = 0.05$. As correlations in the model time series decrease, KS-statistics converge towards those for the i.i.d. variates. (a) As opposed to Fig. 1, here the total number of samples from the model was held constant for each $\Delta t$ at $10^4$ (red), $10^5$ (green) and $10^6$ (blue) analysed control errors. For example, in order to get $10^6$ control errors that occurred 10s separated from each other, $8.5 \cdot 10^8$ control errors were simulated for each of the $10^3$ model time-series to compensate for the subsampling. To only investigate the distribution tails, cutoffs for the Hill estimator were set to use only the largest half of the log-range of the ccdf. For example, for sets consisting of $10^6$ events, only the largest $10^3$ ones were used. (b): The realistic case, were only $10^6$ control errors were simulated for each time series thus reducing the number of analyzed events for increasing $\Delta t$. Here, we chose the cutoff rank by minimizing the KS statistics of the fit.
FIG. 3: Probabilities, that the KS-statistics for power-law fits for the continuous model as in Fig. 1, Fig. 2 are worse than for experimental time series. Triangles indicate comparisons between complete time series and circles time-series where only one event every 10s was used. Cutoffs were again chosen to minimize KS-D-values. Here, we also included a second data set (3b) of subject three, that was recorded a month after the first one (3a). Hence, the subject was not naive anymore. This second data set was not included in the analysis for the main paper. As expected if control errors for model and experimental data were distributed equally, goodness of fits scatter around 0.5.

(a) Combined time series from all trials for each subject. Control errors used for fitting spanned $1.2 \pm 0.3$ orders of magnitudes (oom) before and $0.8 \pm 0.2$ oom after subsampling. Pdf exponents $\delta$ were $3.8 \pm 0.4$ before and $3.5 \pm 0.6$ after subsampling (all values are mean ± standard deviation).

(b) Combined time series of the trials of each day separately for each subject. Control errors used for fitting spanned $1.0 \pm 0.4$ oom before and $0.7 \pm 0.2$ oom after subsampling. Pdf exponents $\delta$ were $3.7 \pm 0.5$ before and $3.4 \pm 0.7$ after subsampling. Forcing the algorithm to use more data points yields on average higher p-values, but less convincing fits.
FIG. 4: Distributions and fits for the combined time series for each subject. Cutoffs are indicated by horizontal lines. $p_\succ$-values are also shown in Fig. 3 (a). The orders of magnitudes of control errors spanned by the data points above each cutoff is indicated by “oom”. $\delta$ again indicates pdf-tail exponents. Dashed lines: Half-normal distributions with the same variance as the respective time series. Time series $3a^*$ is the same as $3a$, except that the very first trial (3 minutes) of this subject has been discarded.
FIG. 5: (color online) (a) Probability according to the KS-statistics, that the tails of two simulations of the continuous control model follow the same distribution. Parameters were again chosen as in Fig. 1. For each data point, $10^2$ tuples of model time-series were compared to each other. As in Fig. 2, the total number of samples from the model was held constant for each $\Delta t$ at $10^4$ (red), $10^5$ (green) and $10^6$ (blue) analysed control errors. Cutoffs were fixed at half of the log-range of the ccdf. Dashed line: $p_{>}$-value of 0.5. Dotted line: $p_{>}$ = 0.05. As correlations in the model time series decrease, KS-statistics shift towards those for the i.i.d. variates, albeit slower than in Fig. 2. (b): Complementary cumulative distribution of balancing errors $|y|$ made by the continuous control model with an exceptionally steep slope (pdf-exponent $\delta = 5.8$) above the range found experimentally. Still, the tail can at least qualitatively be well characterized as a power-law. Parameters were chosen as in Fig. 1 except for $\tau_m = 200\text{ ms}$. This simulation was split into parts that were simulated in parallel. From each one of 100 time series consisting of $10^9$ (after sub-sampling with $\Delta t = 2\text{s}$) events, only the largest $10^7$ ones were kept, combined, and rank ordered again. These events represent the tail of $10^{11}$ events. The part above $F_c = 10^{-4}$ was estimated from a series of $10^7$ time steps, because it shows almost no variations over repeated simulations.
The scaling regimes in the power spectra only extend over a short range on either axis. Even worse, boundary effects extend into both regimes. Therefore, it is not possible to investigate these scaling exponents with the same rigor as the scaling in the error distributions. One should regard the characterization as a broken power law only as a rough description of the observed effects. Also, our basic theory of criticality in adaptive control makes no general claims about spectral densities. However, spectra provide valuable evidence in addition to error distributions allowing to rule out alternative parameter choices or model types. For comparison, see also the similar plot for a discrete time model in supplement 2. While we will investigate the high-frequency scaling for the model (where it can be extended towards higher frequencies) in more details at the end of this section, the first scaling regime below the reaction time is the more interesting one because it corresponds to active control movements.

As shown in Fig. 6 (b), realistic exponents $\lambda_1$ for the scaling regime below 5Hz are found for a combination of cautiously performed control with $\gamma$ slightly above one and fast adaptation with $\tau_m \leq t_r$. Increasing either $\gamma$ or $\tau_m$ increases $\lambda_1$ which can even exceed $\lambda_2$ (not shown). Hence, choosing a long integration time $\tau_m$ destroys the characteristic knee shape of the experimental spectra. However, the power law scaling in the control errors $y$ vanishes as well. Increasing the controllers gain (amplifying control errors) allows to re-establish a power-law tail in $y$, but does not recover the characteristic shape of the power spectrum as shown in Fig. 7. $\gamma$ also determines the position of the low frequency onset of the first scaling regime corresponding to the longest temporal correlation length. Higher gains move it towards higher frequencies. In Fig. 7, this regime has vanished due to the very large gain. For some parameter choices, a peak indicating oscillations may occur.

The high frequency scaling exponent $\lambda_2$ does not depend on model parameters. Close to the knee, small parameter-dependent resonances occur which are reduced when simulations with slightly different parameter sets are combined like in Fig. 1 in the main paper. A similar structure is observed in all experimental spectra. The spectrum levels near the Nyquist frequency.

Fig. 8 shows the spectrum for the continuous model with a box-shaped integration window. This model does generate power-law distributed errors, but the onset of the scaling
regime is less similar to the experimental time series than for the model with exponential forgetting (Fig. 6). The spectrum is also too flat for frequencies between approx. 0.1 and 0.8Hz for integration window lengths $t_r$ that yield realistic error scaling exponents $\delta$. Adjusting $\gamma$ does not lead to more convincing fits.

Fig. 9 shows a non-adaptive controller which is driven into the critical regime by precisely tuned parametric noise. This model does not reproduce experimental spectra. Also, the onset of the scaling in the error distributions is not as convincing as for the adaptive model. Note, that we do not rule out other sources for parametric noise in addition to fast adaptation. However, fast adaptation is the only explanation for why such a noise source should be tuned to a critical point in addition to providing the best fits of experimental results.

Finally, Fig. 10 (a) shows the model simulated with smaller integration steps. Here, it can be clearly seen, that the high-frequency scaling converges towards values of 2 for high frequencies. However, since the experimental time-series do not resolve these high frequencies, we thus far investigated only the onset of this scaling regime. Fig. 10 (b) shows the model driven by strongly lowpass filtered noise. As can be seen, the scaling exponent $\lambda_1$ is at most weakly affected while the high frequency scaling exponent $\lambda_2$ is doubled. This finding underlines, that $\lambda_1$ mainly characterizes the type of active control behavior while $\lambda_2$ characterizes intrinsic noise and passive damping properties of the system. These filtering properties also affect the small resonances close to the reaction time found in the experimental time series.
FIG. 6: Comparison of experimental time series (thin grey lines) and the continuous model (thick black lines) with controller reaction time $t_r = 170\text{ ms}$, gain $\gamma = 1.07$, and memory time constant $\tau_m = 120\text{ ms}$. The controlled systems time constant was $\tau = 250\text{ ms}$ and nose scaling $\sigma = 5\text{ s}^{-0.5}$. Time discretization was $h = 11.8\text{ ms}$ and the simulated time series length was 3.3 hours. Analysis and figure generation methods were identical to Fig. 1 in the main paper. (a): Complementary cumulative distribution $F_c$ of the absolute balancing errors $|y|$. Short diagonal line: a power-law in the pdf $p(y) = |y|^{-\delta}$ corresponds to $F_c(y) = |y|^{-\delta+1}$. Dotted line: Gaussian with the same mean and variance as the simulated time series. (b): Power spectrum.

FIG. 7: Comparison of experimental time series (thin grey lines) and the continuous model with a long memory time constant $\tau_m = 1\text{ s}$ (thick black lines) and high gain $\gamma = 5$. Other parameters, analysis, and figure generation methods were identical to Fig. 6. As opposed to the combination of smaller $\tau_m$ and $\gamma$, here (a) the onset of the power law in the probability distribution and (b) the characteristic knee in the experimental power spectra is not reproduced.
FIG. 8: Comparison of experimental time series (thin grey lines) and the continuous model with a sliding time window for parameter estimation of fixed length $t_m = 0.25\text{ms}$. Compared to previous figures, the overall scale was slightly increased setting $\sigma = 6$ to better fit the experimental data. Still, (a) the onset of the power law in the probability distribution and (b) the power spectra below 1Hz are fitted less convincing than for the model with exponential forgetting (Fig. 6). Other parameters, analysis, and figure generation methods were identical to Fig. 6.
FIG. 9: Comparison of experimental time series (thin grey lines) and the modified model where the estimator $1/\tilde{\vartheta}$ has been replaced by the true system time constant $\tau$ (thick black lines). To obtain power laws, parametric noise has been introduced by choosing the scaling for the driving noise as $\sigma(t) = 0.73 + 0.35 \cdot |\tilde{y}(t)|$. Other parameters, analysis, and figure generation methods were identical to Fig. 6. As opposed to the adaptive model in the main paper, here (a) the onset of the power law in the probability distribution and (b) the characteristic knee in the experimental power spectra is not reproduced. For this model, the shape of the error distribution depends on both, the additive and multiplicative parts of the noise. Therefore, in contrast to the adaptive model, the overall scaling cannot be chosen independently from the tail exponent. Most significantly, the error distribution only resembles a power law for very specific parameter choices. No parameter combinations reproduce the experimental spectra.

FIG. 10: (a): Power spectrum for the continuous model as in Fig. 6, but simulated with a much smaller integration step of $2 \cdot 10^{-4}$s. (b): Power spectrum for the same model as in (a), but driven by lowpass filtered noise with a time-constant of one second.
VARIANCE AND DISPLACEMENT

Multi-scaling in diffusive processes is often investigated using different measures than presented in the main paper. Here, we present two popular measures that allow to compare human virtual balancing behavior to other processes. Fig. 11 (a) shows the dependence of the standard deviation of the series of increments $y(t + \Delta t) - y(t)$. For short increments, the scaling exponent of one characterizes hyper-diffusive behavior (i.e. positive correlations). Around increments of 10 seconds, the scaling approaches the expected value for an uncorrelated random walk of 0.5. Both scaling exponents are always seen for the continuous model and experimental time series, but depend on parameters in time-discrete models [2]. Very similar behavior has been found before for stock-market time series [3].

Fig. 11 (a) shows the dependence of mean square displacements between control errors $y(t), y(t + \Delta t)$ on the time $\Delta t$ by which they are separated. Again, hyperdiffusion is observed for small $\Delta t$. For the smallest timescales, experimental time-series are slightly more correlated than those from the model. Like the power spectra for very high frequencies, this deviation is most likely due to passive damping which is not included in the model. Between $\Delta t$ of the order of the reaction time and approximately 10 seconds, the scaling corresponds to subdiffusion. For even higher $\Delta t$, saturation is observed. This is expected for mean-reverting processes. Similar scaling regimes for displacements have been reported before for center of pressure trajectories of upright standing humans [4, 5]. They have been speculated to represent time scales dominated by open- and closed-loop control. While this idea is basically consistent with the structure of the adaptive control model presented in the main paper, the latter one does not include a mechanism for open-loop control like e.g. damping of joints by reflexes or mechanical properties.
FIG. 11: (a) Scaling of the standard deviation of the cumulated magnitudes and (b) mean displacement for different time intervals $\Delta t$ for the same time series from the continuous model (thick black line) and experiments (thin grey line) as in Fig. 1 in the main Paper. The dashed line indicates the scaling of a random walk ($H = 0.5$).

[6] Note, how this choice of subsampling intervals relates to the correlation structure investigated in Fig. 11.
[7] Minor problems with the described method may still be expected because the exact correlations in the tested time series and the surrogate ones may differ. Also, distributions from model and experimental time series are convex for small $|y|$ and then usually become slightly concave when converging towards an approximately straight line in log-log coordinates. The distributions of the transformed random variates we use for comparison (Fig. 1 (d)) are convex up to the threshold and then exactly follow the power law by construction. However, these potential pitfalls turned out not to be very problematic.
Supplement 2 to “Criticality of adaptive control dynamics”:

Discrete-time limit.

Felix Patzelt and Klaus Pawelzik

Institute for Theoretical Physics, University of Bremen, Germany

(Dated: December 2, 2011)

Abstract

When control and observation in the continuous control model are restricted to discrete times, the whole system can be reduced to a random map. A special case of this discrete-time system is identified: The optimal controller given only the two most recent observations which has already been shown analytically to generate power-law distributed control errors.
THE MINIMAL MODEL

A simple control system with actual dynamics, control and observation taking place only at
discrete times \( t = 1, 2, 3, \ldots \) is given by the random map:

\[
y_{t+1} = \alpha y_t + \beta_t, \tag{1}
\]

with parameter \( \alpha \) and Gaussian distributed independent random variables \( \beta_t \sim \mathcal{N}(0, \sigma) \).

Control by removing a prediction of \( y_{t+1} \) in timestep \( t+1 \) gives

\[
y_{t+1} = (\alpha - \tilde{\alpha}_{t+1}) y_t + \beta_t. \tag{2}
\]

The estimation \( \tilde{\alpha}_{t+1} \) that minimizes the expected error given the two most recent observa-
tions \( \langle y_{t+1}^2 \vert y_t, y_{t-1} \rangle \) is

\[
\tilde{\alpha}_{t+1} = \frac{y_t}{y_{t-1}} + \alpha_t. \tag{3}
\]

We here assumed, that \( \alpha \) can be considered constant over the span of two observations,
but that nothing else is known about its time- or state dependency. Equations (2) and
(3) represent a minimal adaptive control system with a restricted memory. It can also be
written as

\[
y_{t+1} = -\frac{y_t}{y_{t-1}} \beta_{t-1} + \beta_t \tag{4}
\]

and was shown analytically to generate fluctuations with a probability distribution function
(pdf) whose tail obeys a power-law \( p(y) \propto |y|^{-\delta} \) with exponent \( \delta = 2 \) independently of \( \alpha \)
and \( \beta \). Using \( m \geq 2 \) timesteps for the estimator numerically yields \( \delta = m \). [2]

In the following, we demonstrate how the above model can be obtained as a limiting case
of the continuous control model presented in the main paper. Yet, we first recapitulate the
prerequisite model equations for the readers convenience.

THE CONTINUOUS MODEL

The observable dynamics of the continuous control system are defined by

\[
\dot{y}(t) = \frac{1}{\tau} y(t) - \gamma \tilde{\vartheta}(t) \tilde{y}(t) + \beta(t) \tag{5}
\]

with time constant \( \tau \) and Gaussian white noise \( \beta(t) \). The second term on the right hand side
is the controllers contribution which is proportional to the expectation value (prediction)

\[
\tilde{y}(t) = e^{\tilde{\vartheta}(t) t_r} \left( - \int_0^{t_r} e^{-\tilde{\vartheta}(t') t_r} \gamma(t) \tilde{\vartheta}(t - t_r + t') \tilde{y}(t - t_r + t') \, dt' + y(t - t_r) \right). \tag{6}
\]
of $y(t)$ given observations up to time $t - t_r$ where $t_r$ is the controllers reaction time. Here,

$$\tilde{\vartheta}(t + t_r) = \frac{\int_{t-t_m}^{t} y(t') \left(\dot{y}(t') + \gamma \tilde{\vartheta}(t') \tilde{y}(t')\right) dt'}{\int_{t-t_m}^{t} y(t')^2 dt'}.$$  \hfill (7)

is the continuous record maximum likelihood (ML) estimator for $1/\tau$. To make $\dot{y}$ negative on average, the control term contains a gain factor $\gamma > 1$. $t_m$ is the controllers memory length. In the main paper, we modified equation (7) by using an exponentially decaying integration window to obtain a differential formulation. We will omit this step here because because we would have to revert it later on to obtain the desired limit anyway. However, the discrete limit can be performed using the form in the main paper, too.

**PERFORMING THE LIMIT**

Now consider a special case where the controller is not active all the time. Instead of a constant gain, let

$$\gamma \tilde{\vartheta} = \sum_{i=1}^{\infty} \delta(t - i \Delta t), \quad \Delta t = k_r t_r, \quad k_r \in \mathbb{R}$$  \hfill (8)

thereby removing the predicted $y$ completely during short control pulses. In this case, the sde (5) can be solved in between two control pulses (see Fig. 1):

$$y(t + \Delta t - 0) = e^{(\Delta t-0)/\tau} y(t) + \int_{t}^{t+\Delta t-0} \beta(t') dt'$$  \hfill (9)
where the zeroes indicate, that the solution describes the system at time $t + \Delta t$, but before the control pulse at this time has been applied. $y(t)$ which is given includes the control pulse at time $t$. To obtain $y(t + \Delta t)$ after the control pulse has been applied, the latter is simply added. Hence, using the simplified notation

$$y_k = y(k\Delta t)$$
$$k = 1, 2, \ldots$$
$$\alpha = e^{\Delta t/\tau}$$
$$\bar{\alpha} = e^{\Delta t}$$
$$\bar{y}_k = \bar{y}(k\Delta t)$$
$$m = \min\{n \in \mathbb{N} \mid n \geq 2, n \geq t_r/\Delta t\}$$
$$\beta_k = \int_{k\Delta t-1}^{k\Delta t} \beta(t')dt'$$

we obtain

$$y_{k+1} = \alpha y_k - \bar{y}_{k+1} + \beta_k. \quad (10)$$

The prediction (6) can be expressed similarly after inserting (8):

$$\bar{y}_{k+k_r} = \bar{\alpha}_{k+k_r}^k y_k - \sum_{i=1}^{k_r-1} \bar{\alpha}_{k+k_r}^{k_r-i} \bar{y}_{k+i}. \quad (11)$$

Finally, we assume, that the controller only observes the system at the times when the control pulses are applied. We then have to replace the continuous record ML estimator $\hat{\theta}$ for $1/\tau$ (7) by the discrete ML estimator for $\alpha$:

$$\bar{\alpha}_{k+k_r} = \frac{\sum_{i=0}^{m-2} (y_{k-i} + \bar{y}_{k-i}) y_{k-i-1}}{\sum_{i=0}^{m-2} \bar{y}_{k-i-1}^2}. \quad (12)$$

Here, the integral over the observed time interval $[t - t_m, t]$ has been replaced by the sum over the past $m$ observations. Detailed information on this limit can be found e.g. in [1]. For an intuitive understanding, note the big bracket in the numerator in Eq. (7) containing the observed velocity corrected by the controllers action at the time of observation. This term is changed analogously to (10) since interaction between controller and system now takes place only during control pulses. In (12), the bracket in the numerator therefore contains the observed $y$ corrected by the control pulse at that time.
Eqns. (10) - (12) represent the discrete-time limit of the continuous model Eqns. (5) - (7) under the condition (8) that the controller interacts with the system only during short control pulses. Choosing the reaction delay \( k_r = 1 \) and memory \( m = 2 \) (therefore dropping the sums), the discrete system reduces to the minimal model Eqns. (2), (3) (renaming \( k \) into \( t \)).

**DISCUSSION**

We performed the discrete time limit such, that during each pulse the expectation value of \( y \) is removed completely. Hence, the controller is optimal given the pulse times and memory length. However, this controller is not able to faithfully reproduce the experimental findings presented in the main paper (see below). These findings require a low control gain, reducing the predicted \( y \) very carefully. Adding a gain to the discrete case severely hampers the controller and does not allow to reproduce the experimental findings (not shown). It is also not possible to relate the discrete time steps to any time scale in human behavior because their behavior bears no evidence for the use of similar control pulses. Results similar to the experimental data were obtained earlier by introducing a delay for the estimator, but not introducing a real reaction time [2]. This creates two scaling regimes which are qualitatively similar to the data, but quantitatively have the wrong scaling exponents in the spectra and the wrong Hurst exponents (see supplement one). A real reaction time in the discrete case as in the above model creates very pronounced resonances that are not observed in the experimental time series (Fig. 2 (b)). These strong resonances also occur, when an estimator with exponentially decaying weights is used (not shown). Control errors are still power-law distributed as shown in Fig. 2 (a). For a minimal reaction delay of one timestep, using \( m \geq 2 \) timesteps for the estimator numerically yields a pdf tail exponent \( \delta = m \), independent of the parameters \( \alpha \) and \( \beta \). Longer reaction delays decrease \( \delta \). This widening of the distribution tails is stronger for small values of \( \alpha \), breaking the invariance against the systems time constant.
FIG. 2: (a) Cumulated complementary distribution function for the discrete model defined by Eqns. (10), (11) and (12) with parameter $\alpha = 2$; memory length $m = 4$ steps, reaction delay $k_r = 10$ steps, noise level $\sigma = 1$. Dashed line: A Gaussian with the same variance as the model time series.

(b): Power spectrum for the same time series. Both subfigures were created like Fig. 1 (a), (b) in the main paper.
