For a Brownian dynamics system with a 2d reaction coordinate, the AWH detection of covering gets delayed because the check frequency is determined to be too high. The check frequency is set to scale with the number of points in the grid but the covering time should not directly depend on the discretization. For Brownian dynamics, the time step can be chosen to be large relative to the diffusion time. This exposed the problem since the number of samples needed per covering time is relatively small.

The same step interval was also used for checking for histogram anomalies, which however only generates warnings and is not integral to the AWH method itself.

Now the covering check interval is instead determined from the number of sigmas, the “width” of one sample, required to cover each dimension. Since there is about 1 point per sigma, the dependency of the check interval on the number of points is essentially unchanged in the one-dimensional case. However, this relation is not numerically exact. Therefore, this change also requires updating reference data for a regression test to be updated.

The check interval for histogram anomalies is set to be the same as for the covering, but could in the future be made less frequent.

Added release note.

Fixes #2487

Change-Id: I5af0d48436664d8fdfe8bafa05cde3cdae27e45a
Looking at the threshold value for covering, I think the least amount of samples you need to cover \( \#\text{points}_{in\_dim}/\sqrt{2 \pi} \). Or did I miss something? So the fastest you can cover is in \( 0.4 \#\text{points}_{in\_dim} \). So we check somewhat infrequent for the fastest possible covering case. But it's rather unlikely that diffusion covers the range homogeneously. The time should go up with \( \sqrt{\#\text{points}} \).

The covering frequency does not depend on discretization. The discretization is set by the force constant. So the covering frequency depends on force constant, but that makes sense, I think, because this changes the requested resolution.

There's a negation missing somewhere? The covering frequency should not depend on either force constant or discretization (the latter two are functions of each other).

Looking at the threshold value for covering, I think the least amount of samples you need to cover \( \#\text{points}_{in\_dim}/\sqrt{2 \pi} \). Or did I miss something? So the fastest you can cover is in \( 0.4 \#\text{points}_{in\_dim} \). So we check somewhat infrequent for the fastest possible covering case. But it's rather unlikely that diffusion covers the range homogeneously. The time should go up with \( \sqrt{\#\text{points}} \).

The covering criterion should AFAIR not depend on the \# of points. The weight required for each point will decrease with the number of points. Correct me if I'm wrong...

I did realize however that the problem is there are two quite different checks being determined by the isCheckStep function. I'm adding another change that splits this into two: one for covering and one for histogram anomalies. After looking closer the latter actually should depend on the number of points, since there the realized visits histogram, which is a simple binning histogram, is being compared to the weight histogram (to see e.g. if the force constant may be too low).

I don't see how the threshold depends on the number of points. It looks like the prefactor in the Gaussian to me.

I don't see how the threshold depends on the number of points. It looks like the prefactor in the Gaussian to me.

Exactly, it should not depend on the number of points in the end

\[
\text{weightThreshold} \ast= \text{grid.axis(d).spacing()} \ast \text{std::sqrt(dimParams[d].betak \ast 0.5 \ast M_\_1 PI)};
\]

the spacing will decrease with increasing number of points.

so the threshold is fixed at 0.4 and does not change with the number of points.

so the threshold is fixed at 0.4 and does not change with the number of points.

You are right. The confusion on my part comes from that the transition of the coordinate should not depend on discretization, but because we don't sample every step and the covering criterion checks for covering each point rather than transitions, there's a built in dependency on the number of points.

You are right. The confusion on my part comes from that the transition of the coordinate should not depend on discretization, but because we don't sample every step and the covering criterion checks for covering each point rather than transitions, there's a built in dependency on the number of points.

Applied in changeset 26db2ac4a3a0e5d6e3be086151c1e6bde47cc1a1.
Is this issue resolved by the recently submitted changes?

#11 - 06/14/2018 10:50 AM - Mark Abraham
- Status changed from Resolved to Closed

Files

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>bd-2d.tgz</td>
<td>2.91 KB</td>
<td>04/25/2018</td>
<td>Viveca Lindahl</td>
</tr>
</tbody>
</table>