Journal of Machine Learning Research 11 (2010) 1-9 Submitted 11/10; Published 7/11

Erratum: A Generalized Path Integral Control Approach to Reinforcement Learning

Evangelos A.Theodorou ETHEODOR@USC.EDU
Jonas Buchli JONAS@BUCHLI.ORG
Stefan Schaal* SSCHAAL@USC.EDU
Department of Computer Science
University of Southern California
Los Angeles, CA 90089-2905, USA

Editor: Daniel Lee

In this erratum we correct a mistake in the derivation of the generalized path integral control in lemma 2. More precisely, we show that the term **b** in Equation (20) should not appear at all. This mistake does not affect any of the results presented in this paper, as the **b** term always dropped out in all of our applications.

The changes are:[1]

1. Equation $Z(\tau_i) = \tilde{S}(\tau_i) + \frac{\lambda(N-i)l}{2}\log(2\pi dt)$, in page (3144) should change to:

$$Z(\tau_i) = \tilde{S}(\tau_i) + \frac{\lambda(N-i)l}{2}\log\left(2\pi\sigma_\varepsilon^2 dt\right),$$

   where σ_ε^2 is defined such that $\Sigma_\varepsilon = \sigma_\varepsilon^2 I$ and the path dependent cost $\tilde{S}(\tau_i)$ is defined:

$$\tilde{S}(\tau_i) = S(\tau_i) + \frac{\lambda}{2}\sum_{j=i}^{N-1}\log|\mathcal{B}(\mathbf{x}_{t_j}, t_j)|,$$

   with $\mathcal{B}(\mathbf{x}_{t_j}, t_j) = \mathbf{G}(\mathbf{x}, t_j)\mathbf{G}(\mathbf{x}, t_j)^T$.

2. Equation $\lim_{dt\to 0}\left(\nabla_{\mathbf{x}_{t_i}^{(c)}}\tilde{S}(\tau_i)\right) = -\mathbf{H}_{t_i}^{-1}\left(\mathbf{G}_{t_i}^{(c)}\varepsilon_{t_i} - \mathbf{b}_{t_i}\right)$, on page (3144) should change according to the lemma 2 as it is given in this erratum.

3. Equation (20) on page (3145) should change to:

$$\mathbf{u}_L(\tau_i) = \mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)T}\left(\mathbf{G}_{t_i}^{(c)}\mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)T}\right)^{-1}\mathbf{G}_{t_i}^{(c)}\varepsilon_{t_i}.$$

4. Equation $\mathbf{u}_L(\tau_i) = \frac{\mathbf{R}^{-1}\mathbf{g}_{t_i}^{(c)}}{\mathbf{g}_{t_i}^{(c)T}\mathbf{R}^{-1}\mathbf{g}_{t_i}^{(c)}}\left(\mathbf{g}_{t_i}^{(c)T}\varepsilon_{t_i} - \mathbf{b}_{t_i}\right)$, on page (3145) should change to

∗. Also at ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan.

1. An updated version of the paper which includes this erratum can be found at `http://www-clmc.usc.edu/Resources/Publications?id=10413`.

$$\mathbf{u}_L(\tau_i) = \frac{\mathbf{R}^{-1}\mathbf{g}_{t_i}^{(c)}}{\mathbf{g}_{t_i}^{(c)T}\mathbf{R}^{-1}\mathbf{g}_{t_i}^{(c)}}\mathbf{g}_{t_i}^{(c)T}\varepsilon_{t_i}.$$

5. Equation (21) on page (3147) should change to $\mathbf{u}_L(\tau_i) = \varepsilon_{t_i}$.

6. Equation $\mathbf{u}_L(\tau_i) = \varepsilon_{t_i} - \mathbf{G}_{t_i}^{-1}\mathbf{b}_{t_i}$, on page (3147), should change to $\mathbf{u}_L(\tau_i) = \varepsilon_{t_i}$.

Next we provide the updated proofs for appendix of the initial paper.

Acknowledgments

We are grateful to Bert Kappen for providing his insightful feedback on our initial result.

Appendix A.

Lemma 1 ***The optimal control solution to the stochastic optimal control problem expressed by (1),(2),(3) and (4) is formulated as:***

$$\mathbf{u}_{t_i} = \lim_{dt\to 0}\left[-\mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)T}\int \tilde{p}(\tau_i)\nabla_{\mathbf{x}_{t_i}^{(c)}}\tilde{S}(\tau_i)d\tau_i\right],$$

where $\tilde{p}(\tau_i) = \frac{\exp\left(-\frac{1}{\lambda}\tilde{S}(\tau_i)\right)}{\int\exp\left(-\frac{1}{\lambda}\tilde{S}(\tau_i)\right)d\tau_i}$ ***is a path dependent probability distribution. The term*** $\tilde{S}(\tau_i)$ ***is a path function defined as*** $\tilde{S}(\tau_i) = S(\tau_i) + \frac{\lambda}{2}\sum_{j=i}^{N-1}\log|\mathcal{B}(\mathbf{x}_{t_j}, t_j)|$ ***that satisfies the following condition*** $\lim_{dt\to 0}\int\exp\left(-\frac{1}{\lambda}\tilde{S}(\tau_i)\right)d\tau_i \in \mathcal{C}^{(1)}$ ***for any sampled trajectory starting from state*** $\mathbf{x}_{t_i}$***. Moreover the term*** $\mathbf{H}_{t_j}$ ***is given by*** $\mathbf{H}_{t_j} = \mathbf{G}_{t_j}^{(c)}\mathbf{R}^{-1}\mathbf{G}_{t_j}^{(c)T}$ ***while the term*** $S(\tau_i)$ ***is defined according to***

$$S(\tau_i) = \phi_{t_N} + \sum_{j=i}^{N-1} q_{t_j}dt + \frac{1}{2}\sum_{j=i}^{N-1}\left\|\frac{\mathbf{x}_{t_{j+1}}^{(c)} - \mathbf{x}_{t_j}^{(c)}}{dt} - \mathbf{f}_{t_j}^{(c)}\right\|_{\mathbf{H}_{t_j}}^2 dt.$$

Proof The proof is exactly the same with the one of the initial manuscript. ■

Lemma 2 ***Given the stochastic dynamics and the cost in (1),(2),(3) and (4) the gradient of the path function*** $\tilde{S}(\tau_i)$***, with respect to the directly actuated part of the state*** $\mathbf{x}_{t_i}^{(c)}$ ***is formulated as:***

$$\nabla_{\mathbf{x}_{t_i}^{(c)}}\tilde{S}(\tau_i) = \frac{1}{2dt}\alpha_{t_i}^T\left(\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{H}_{t_i}^{-1}\right)\alpha_{t_i} - \mathbf{H}_{t_i}^{-1}\left(\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}\right)\alpha_{t_i} - \frac{1}{dt}\mathbf{H}_{t_i}^{-1}\alpha_{t_i} + \frac{\lambda}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\log|\mathcal{B}_{t_i}|,$$

where $\mathbf{H}_{t_i} = \mathbf{G}_{t_i}^{(c)}\mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)^T}$**,** $\mathcal{B}_{t_i} = \mathbf{G}_{t_i}^{(c)}\mathbf{G}_{t_i}^{(c)^T}$ ***and*** $\alpha_{t_j} = \left(\mathbf{x}_{t_{i+1}}^{(c)} - \mathbf{x}_{t_i}^{(c)} - \mathbf{f}_{t_i}^{(c)}dt\right)$**.**

Proof:

We are calculating the term $\nabla_{\mathbf{x}_{t_o}^{(c)}} \tilde{S}(\tau_o)$. More precisely we have shown that

$$\tilde{S}(\tau_i) = \phi_{t_N} + \sum_{j=i}^{N-1} q_{t_j} dt + \frac{1}{2} \sum_{j=i}^{N-1} \| \frac{\mathbf{x}_{t_{j+1}}^{(c)} - \mathbf{x}_{t_j}^{(c)}}{dt} - \mathbf{f}_{t_j}^{(c)} \|_{\mathbf{H}_{t_j}}^2 dt + \frac{\lambda}{2} \sum_{j=i}^{N-1} \log |\mathcal{B}_{t_j}|.$$

To limit the length of our derivation we introduce the notation $\gamma_{t_j} = \alpha_{t_j}^T \ \mathbf{h}_{t_j}^{-1} \ \alpha_{t_j}$ and $\alpha_{t_j} = \left(\mathbf{x}_{t_{j+1}}^{(c)} - \mathbf{x}_{t_j}^{(c)} - \mathbf{f}_{t_j}^{(c)} dt \right)$ and it is easy to show that $\| \frac{\mathbf{x}_{t_{j+1}}^{(c)} - \mathbf{x}_{t_j}^{(c)}}{dt} - \mathbf{f}_{t_j}^{(c)} \|_{\mathbf{H}_{t_j}}^2 dt = \frac{1}{dt} \gamma_{t_j}$ and therefore we will have:

$$\tilde{S}(\tau_i) = \phi_{t_N} + \frac{1}{2dt} \sum_{j=i}^{N-1} \gamma_{t_j} + \sum_{t_o}^{t_N} Q_{t_j} dt + \frac{\lambda}{2} \sum_{j=i}^{N-1} \log |\mathcal{B}_{t_j}|.$$

In the analysis that follows we provide the derivative of the 1th, 2th and 4th term of the cost function. For simplicity we will assume that the cost of the state during the time horizon $Q_{t_i} = 0$. In cases that this is not true then the derivative $\nabla_{\mathbf{x}_{t_i}^{(c)}} \sum_{t_i}^{t_N} Q_{t_i} dt$ results in 0. By calculating the term $\nabla_{\mathbf{x}_{t_o}^{(c)}} \tilde{S}(\tau_o)$ we can find the local controls $\mathbf{u}(\tau_i)$. It is important to mention that the derivative of the path cost $S(\tau_i)$ is taken only with respect to the current state $\mathbf{x}_{t_i}$.

The first term is:

$$\nabla_{\mathbf{x}_{t_i}^{(c)}} (\phi_{t_N}) = 0.$$

A.1 Derivative of the 2th Term $\nabla_{\mathbf{x}_{t_i}^{(c)}} \left[\frac{1}{2dt} \sum_{i=1}^{N-1} \gamma_{t_i} \right]$ of the cost $S(\tau_i)$.

The second term can be found as follows:

$$\nabla_{\mathbf{x}_{t_i}^{(c)}} \left[\frac{1}{2dt} \sum_{j=i}^{N-1} \gamma_{t_j} \right].$$

The operator $\nabla_{\mathbf{x}_{t_o}^{(c)}}$ is linear and it can massaged inside the sum:

$$\frac{1}{2dt} \sum_{j=i}^{N-1} \nabla_{\mathbf{x}_{t_j}^{(c)}} \left(\gamma_{t_j} \right).$$

Terms that do not depend on $\mathbf{x}_{t_i}^{(c)}$ drop and thus we will have:

$$\frac{1}{2dt} \nabla_{\mathbf{x}_{t_i}^{(c)}} \gamma_{t_i}.$$

Substitution of the parameter $\gamma_{t_i} = \alpha_{t_i}^T \ \mathbf{H}_{t_i}^{-1} \ \alpha_{t_i}$ will result in:

$$\frac{1}{2dt} \nabla_{\mathbf{x}_{t_i}^{(c)}} \left[\alpha_{t_i}^T \ \mathbf{H}_{t_i}^{-1} \ \alpha_{t_i} \right].$$

By making the substitution $\beta_{t_i} = \mathbf{H}_{t_i}^{-1} \ \alpha_{t_i}$ and applying the rule $\nabla \left(\mathbf{u}(\mathbf{x})^T \mathbf{v}(\mathbf{x}) \right) = \nabla \left(\mathbf{u}(\mathbf{x}) \right) \mathbf{v}(\mathbf{x}) + \nabla \left(\mathbf{v}(\mathbf{x}) \right) \mathbf{u}(\mathbf{x})$ we will have that:

$$\frac{1}{2dt}\left[\nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}\ \ \beta_{t_i}+\nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \ \alpha_{t_i}\right]. \tag{1}$$

Next we find the derivative of α_{t_o}:

$$\nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}=\nabla_{\mathbf{x}_{t_i}^{(c)}}\left[\mathbf{x}_{t_{i+1}}^{(c)}-\mathbf{x}_{t_i}^{(c)}-\mathbf{f}_c(\mathbf{x}_{t_i})dt\right].$$

and the result is

$$\nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}=-I_{l\times l}-\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}dt.$$

We substitute back to (1) and we will have:

$$\frac{1}{2dt}\left[-\left(I_{l\times l}+\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}dt\right)\ \ \beta_{t_i}+\nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \ \alpha_{t_i}\right].$$

$$-\frac{1}{2dt}\left(I_{l\times l}+\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}dt\right)\ \ \beta_{t_i}+\frac{1}{2dt}\ \ \nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \alpha_{t_i}.$$

After some algebra the result of $\nabla_{\mathbf{x}_{t_i}^{(c)}}\left(\frac{1}{2dt}\sum_{j=i}^{N-1}\gamma_{t_i}\right)$ is expressed as:

$$-\frac{1}{2dt}\ \ \beta_{t_i}-\frac{1}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}\ \ \beta_{t_i}+\frac{1}{2dt}\ \nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \ \alpha_{t_i}.$$

A.2 First Subterm: $-\frac{1}{2dt}\ \beta_{t_i}$

We will continue our analysis by finding the limit for each one of the 3 terms above. The limit of the first term is calculated as follows:

$$\begin{aligned}\left(-\frac{1}{2dt}\ \ \beta_{t_i}\right)&=-\left(\frac{1}{2dt}\ \ \mathbf{H}_{t_i}^{-1}\ \ \alpha_{t_i}\right)\\&=-\frac{1}{2}\ \ \mathbf{H}_{t_i}^{-1}\ \ \alpha_{t_i}\\&=-\frac{1}{2}\ \ \mathbf{H}_{t_i}^{-1}\ \ \left((\mathbf{x}_{t_{i+1}}^{(c)}-\mathbf{x}_{t_i}^{(c)})\frac{1}{dt}-\mathbf{f}_{t_i}^{(c)}\right).\end{aligned}$$

A.3 Second Subterm: $-\frac{1}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}\ \ \beta_{t_i}$

The limit of the second term is calculated as follows:

$$\begin{aligned}\left(\frac{1}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}\ \ \beta_{t_i}\right)&=-\frac{1}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_c(\mathbf{x}_{t_i})\ \ \beta_{t_i}=-\frac{1}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_{t_i}^{(c)}\ \ \left(\mathbf{H}_{t_i}^{-1}\ \ \alpha_{t_i}\right)\\&=-\frac{1}{2}\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{f}_c(\mathbf{x}_{t_i})\ \ \mathbf{H}_{t_i}^{-1}\ \ \alpha_{t_i}.\end{aligned}$$

A.4 Third Subterm: $\frac{1}{2dt}\ \nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \alpha_{t_i}$

Finally the limit of the third term can be found as:

$$\left(\frac{1}{2dt}\ \nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \alpha_{t_i}\right) = \nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \left(\frac{1}{2dt}\ \alpha_{t_i}\right) = \nabla_{\mathbf{x}_{t_i}^{(c)}}\beta_{t_i}\ \frac{1}{2}\left((\mathbf{x}_{t_{i+1}}^{(c)} - \mathbf{x}_{t_i}^{(c)})\frac{1}{dt} - \mathbf{f}_{t_i}^{(c)}\right).$$

We substitute $\beta_{t_i} = \mathbf{H}_{t_i}^{-1}\ \alpha_{t_i}$ and write the matrix $\mathbf{H}_{t_i}^{-1}$ in row form:

$$= \nabla_{\mathbf{x}_{t_i}^{(c)}}\left(\mathbf{H}_{t_i}^{-1}\ \alpha_{t_i}\right)\ \frac{1}{2dt}\alpha_{t_i} =$$
$$= \nabla_{\mathbf{x}_{t_i}^{(c)}}\left(\begin{bmatrix} \mathbf{H}_{t_i}^{(1)^{-T}} \\ \mathbf{H}_{t_i}^{(2)^{-T}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{H}_{t_i}^{(l)^{-T}} \end{bmatrix}\ \alpha_{t_i}\right)\ \frac{1}{2dt}\alpha_{t_i} = \nabla_{\mathbf{x}_{t_i}^{(c)}}\begin{bmatrix} \mathbf{H}_{t_i}^{(1)^{-T}}\ \alpha_{t_i} \\ \mathbf{H}_{t_i}^{(2)^{-T}}\ \alpha_{t_i} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{H}_{t_i}^{(l)^{-T}}\ \alpha_{t_i} \end{bmatrix}\frac{1}{2dt}\alpha_{t_i}.$$

We can push the operator $\nabla_{\mathbf{x}_{t_i}^{(c)}}$ inside the matrix and apply it to each element.

$$= \begin{bmatrix} \nabla^T_{\mathbf{x}_{t_i}^{(c)}}\left(\mathbf{H}_{t_i}^{(1)^{-T}}\ \alpha_{t_i}\right) \\ \nabla^T_{\mathbf{x}_{t_i}^{(c)}}\left(\mathbf{H}_{t_i}^{(2)^{-T}}\ \alpha_{t_i}\right) \\ \cdot \\ \cdot \\ \cdot \\ \nabla^T_{\mathbf{x}_{t_i}^{(c)}}\left(\mathbf{H}_{t_i}^{(l)^{-T}}\ \alpha_{t_i}\right) \end{bmatrix}\ \frac{1}{2dt}\alpha_{t_i}.$$

We again use the rule $\nabla\left(\mathbf{u}(\mathbf{x})^T\mathbf{v}(\mathbf{x})\right) = \nabla\left(\mathbf{u}(\mathbf{x})\right)\mathbf{v}(\mathbf{x}) + \nabla\left(\mathbf{v}(\mathbf{x})\right)\mathbf{u}(\mathbf{x})$ and thus we will have:

$$= \begin{bmatrix} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{H}_{t_i}^{(1)^{-T}}\ \alpha_{t_i} + \nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}\ \mathbf{H}_{t_i}^{(1)^{-T}}\right)^T \\ \left(\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{H}_{t_i}^{(2)^{-T}}\ \alpha_{t_i} + \nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}\ \mathbf{H}_{t_i}^{(2)^{-T}}\right)^T \\ \cdot \\ \cdot \\ \cdot \\ \left(\nabla_{\mathbf{x}_{t_i}^{(c)}}\mathbf{H}_{t_i}^{(l)^{-T}}\ \alpha_{t_i} + \nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}\ \mathbf{H}_{t_i}^{(l)^{-T}}\right)^T \end{bmatrix}\frac{1}{2dt}\alpha_{t_i}.$$

We can split the matrix above into two terms and then we pull out the terms α_{t_i} and $\nabla_{\mathbf{x}_{t_i}^{(c)}}\alpha_{t_i}$ respectively :

$$
\begin{aligned}
&= \left(\alpha_{t_i}^T \begin{bmatrix} \nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{(1)^{-T}} \\ \nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{(2)^{-T}} \\ \cdot \\ \cdot \\ \cdot \\ \nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{(l)^{-T}} \end{bmatrix} + \begin{bmatrix} \mathbf{H}_{t_i}^{(1)^{-T}} \\ \mathbf{H}_{t_i}^{(2)^{-T}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{H}_{t_i}^{(l)^{-T}} \end{bmatrix} \nabla_{\mathbf{x}_{t_i}^{(c)}} \alpha_{t_i}^T \right) \frac{1}{2dt} \alpha_{t_i} \\
&= \left(\alpha_{t_i}^T \ \nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} + \mathbf{H}_{t_i}^{-1} \ \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \alpha_{t_i}^T \right) \right) \frac{1}{2dt} \alpha_{t_i}. \\
&= \frac{1}{2dt} \left(\alpha_{t_i}^T \ \nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \alpha_{t_i} + \mathbf{H}_{t_i}^{-1} \ \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \alpha_{t_i}^T \right) \alpha_{t_i} \right).
\end{aligned}
$$

Since $\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \alpha_{t_i}^T \right) = -I_{l \times l} - \nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{f}_{t_i}^{(c)} dt$. the final result is expressed as follows

$$
\frac{1}{2dt} \nabla_{\mathbf{x}_{t_i}^{(c)}} \beta_{t_i} \ \alpha_{t_i} = \frac{1}{2dt} \left[\alpha_{t_i}^T \ \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \ \alpha_{t_i} - \mathbf{H}_{t_i}^{-1} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{f}_{t_i}^{(c)} \right) dt \alpha_{t_i} - \mathbf{H}_{t_i}^{-1} \alpha_{t_i} \right].
$$

A.5 Derivative of the Fourth Term $\nabla_{\mathbf{x}_{t_i}^{(c)}} \left(\frac{\lambda}{2} \sum_{j=i}^{N-1} \log |\mathcal{B}_{t_j}| \right)$ of the cost $S(\tau_i)$.

The analysis for the 4th term is given below:

$$
\nabla_{\mathbf{x}_{t_i}^{(c)}} \left(\frac{\lambda}{2} \sum_{j=i}^{N-1} \log |\mathcal{B}_{t_j}| \right) = \frac{\lambda}{2} \nabla_{\mathbf{x}_{t_i}^{(c)}} \log |\mathcal{B}_{t_i}|.
$$

$$
\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) = \frac{1}{2dt} \alpha_{t_i}^T \ \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \ \alpha_{t_i} - \mathbf{H}_{t_i}^{-1} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{f}_{t_i}^{(c)} \right) \alpha_{t_i} - \frac{1}{dt} \mathbf{H}_{t_i}^{-1} \alpha_{t_i} + \frac{\lambda}{2} \nabla_{\mathbf{x}_{t_i}^{(c)}} \log |\mathcal{B}_{t_i}|.
$$

Theorem 3 ***The optimal control solution to the stochastic optimal control problem expressed by (1),(2), (3) and (4) is formulated by the equation that follows:***

$$
\mathbf{u}_{t_i} dt = \int \tilde{p}(\tau_i) \ \mathbf{u}_L(\tau_i) \ d\tau_i,
$$

where $\tilde{p}(\tau_i) = \frac{\exp\left(-\frac{1}{\lambda} \tilde{S}(\tau_i)\right)}{\int \exp\left(-\frac{1}{\lambda} \tilde{S}(\tau_i)\right) d\tau_i}$ ***is a path depended probability distribution and the term*** $\mathbf{u}(\tau_i)$ ***defined as*** $\mathbf{u}_L(\tau_i) = \mathbf{R}^{-1} \mathbf{G}_{t_i}^{(c)T} \left(\mathbf{G}_{t_i}^{(c)} \mathbf{R}^{-1} \mathbf{G}_{t_i}^{(c)T} \right)^{-1} \mathbf{G}_{t_i}^{(c)} \varepsilon_{t_i}$***, are the local controls of each sampled trajectory starting from state*** $\mathbf{x}_{t_i}$***. The term is defined as*** $\mathbf{H}_{t_i} = \mathbf{G}_{t_i}^{(c)} \mathbf{R}^{-1} \mathbf{G}_{t_i}^{(c)^T}$.

Proof :

To prove the theorem we make use of the lemma **L2** and we substitute $\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i)$ in the main result of lemma **L1**. More precisely from lemma **L1** we have that:

$$\mathbf{u}_{t_i} dt = -\mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)T} dt \int \tilde{p}(\tau_i) \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) \right) d\tau_i.$$

$$\begin{aligned}\mathbf{u}_{t_i} dt &= -\mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)T} dt \int \tilde{p}(\tau_i) \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) \right) d\tau_i \\ &= \mathbf{R}^{-1}\mathbf{G}_{t_i}^{(c)T} E_{\tilde{p}(\tau_i)} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) dt \right).\end{aligned}$$

Now we will find the term $E_{\tilde{p}(\tau_i)} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) dt \right)$. More precisely we will have that:

$$\begin{aligned}E_{\tilde{p}(\tau_i)} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) dt \right) &= E_{\tilde{p}(\tau_i)} \left(\frac{1}{2} \alpha_{t_i}^T \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \alpha_{t_i} \right) - E_{\tilde{p}(\tau_i)} \left(\mathbf{H}_{t_i}^{-1} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{f}_{t_i}^{(c)} \right) \alpha_{t_i} dt \right) \\ &\quad - E_{\tilde{p}(\tau_i)} \left(\mathbf{H}_{t_i}^{-1} \alpha_{t_i} \right) + E_{\tilde{p}(\tau_i)} \left(\frac{\lambda}{2} \nabla_{\mathbf{x}_{t_i}^{(c)}} \log |\mathcal{B}_{t_i}| dt \right).\end{aligned}$$

The first term of the expectation above is calculated as follows:

$$\begin{aligned}E_{\tilde{p}(\tau_i)} \left(\frac{1}{2dt} \alpha_{t_i}^T \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \alpha_{t_i} \right) &= E_{\tilde{p}(\tau_i)} \left(\frac{1}{2} \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \alpha_{t_i} \alpha_{t_i}^T \right) \\ &= \frac{1}{2} E_{\tilde{p}(\tau_i)} \left(trace \left(\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \alpha_{t_i} \alpha_{t_i}^T \right) \right) \\ &= \frac{1}{2} trace \left(\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) E_{\tilde{p}(\tau_i)} \left(\alpha_{t_i} \alpha_{t_i}^T \right) \right).\end{aligned}$$

By taking into account the fact that $E_{\tilde{p}(\tau_i)} \left(\alpha_{t_i} \alpha_{t_i}^T \right) = \mathbf{G}_{t_i}^{(c)} \Sigma_{\varepsilon} \mathbf{G}_{t_i}^{(c)^T} dt$

$$\begin{aligned}E_{\tilde{p}(\tau_i)} \left(\frac{1}{2dt} \alpha_{t_i}^T \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \alpha_{t_i} \right) &= \frac{1}{2} trace \left(\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \mathbf{G}_{t_i}^{(c)} \Sigma_{\varepsilon} \mathbf{G}_{t_i}^{(c)^T} dt \right) \\ &= \frac{dt}{2} trace \left(\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1} \right) \mathbf{G}_{t_i}^{(c)} \Sigma_{\varepsilon} \mathbf{G}_{t_i}^{(c)^T} \right).\end{aligned}$$

By using the fact that the noise and the controls are related via $\Sigma_{t_j} = \mathbf{G}_{t_j}^{(c)} \Sigma_{\varepsilon} \mathbf{G}_{t_j}^{(c)^T} dt = \lambda \mathbf{G}_{t_j}^{(c)} \mathbf{R}^{-1} \mathbf{G}_{t_j}^{(c)^T} dt = \lambda \mathbf{H}_{t_j} dt$ with $\mathbf{H}_{t_j} = \mathbf{G}_{t_j}^{(c)} \mathbf{R}^{-1} \mathbf{G}_{t_j}^{(c)^T}$ we will have:

$$\begin{aligned} E_{\tilde{p}(\tau_i)}\left(\frac{1}{2}\alpha_{t_i}^T \left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1}\right) \alpha_{t_i}\right) &= \frac{1}{2} trace\left(\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{H}_{t_i}^{-1}\right) \mathbf{G}_{t_i}^{(c)} \Sigma_{\varepsilon} \mathbf{G}_{t_i}^{(c)^T}\right) \\ &= \frac{\lambda dt}{2} trace\left(\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathcal{B}(\mathbf{x}_{t_i})^{-1}\right) \mathcal{B}(\mathbf{x}_{t_i})\right) \\ &= \frac{\lambda dt}{2} \nabla_{\mathbf{x}_{t_i}^{(c)}} \log |\mathcal{B}(\mathbf{x}_{t_i})|^{-1} \\ &= -\frac{\lambda dt}{2} \nabla_{\mathbf{x}_{t_i}^{(c)}} \log |\mathcal{B}(\mathbf{x}_{t_i})|. \end{aligned}$$

The second term $E_{\tilde{p}(\tau_i)}\left(\mathbf{H}_{t_i}^{-1}\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{f}_{t_i}^{(c)}\right) \alpha_{t_i} dt\right) = 0$ since $dt\alpha_{t_i} = dt\mathbf{G}_{t_i}^{(c)} d\mathbf{w} \to 0$. It is important here to note that the term α_{t_i} has two equivalent interpretations which are $\alpha_{t_i} = \mathbf{G}_{t_i}^{(c)} L d\mathbf{w}$ where $LL^T = \Sigma_{\varepsilon}$ or $\alpha_{t_i} = \mathbf{G}_{t_i}^{(c)} \sqrt{dt}\varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \Sigma_{\varepsilon})$. The first representation is the more proper one and leads to $dt\alpha_{t_i} = dt\mathbf{G}_{t_i}^{(c)} d\mathbf{w} \to 0$ because $dt d\mathbf{w} \to 0$. With the equation above we will have that:

$$\begin{aligned} E_{\tilde{p}(\tau_i)}\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \tilde{S}(\tau_i) dt\right) &= -E_{\tilde{p}(\tau_i)}\left(\mathbf{H}_{t_i}^{-1}\left(\nabla_{\mathbf{x}_{t_i}^{(c)}} \mathbf{f}_{t_i}^{(c)}\right) \alpha_{t_i} dt\right) - E_{\tilde{p}(\tau_i)}\left(\mathbf{H}_{t_i}^{-1} \alpha_{t_i}\right) \\ &= -E_{\tilde{p}(\tau_i)}\left(\mathbf{H}_{t_i}^{-1} \alpha_{t_i}\right). \end{aligned}$$

Substituting back to the optimal control we will have that:

$$\mathbf{u}_{t_i} dt = \int \tilde{p}(\tau_i) \mathbf{R}^{-1} \mathbf{G}_{t_i}^{(c)T} \mathbf{H}_{t_i}^{-1} \alpha_{t_i} d\tau_i,$$

$$\mathbf{u}_{t_i} = \int \tilde{p}(\tau_i) \mathbf{R}^{-1} \mathbf{G}_{t_i}^{(c)T} \mathbf{H}_{t_i}^{-1} \mathbf{G}_{t_i}^{(c)} \varepsilon_{t_i} d\tau_i,$$

or in a more compact form:

$$\boxed{\mathbf{u}_{t_i} dt = \int \tilde{p}(\tau_i)\ \mathbf{u}_L^{(dt)}(\tau_i)\ d\tau_i,}$$

where the local controls $\mathbf{u}_L^{(dt)}(\tau_i)$ are given as follows:

$$\mathbf{u}_L^{(dt)}(\tau_i) = \mathbf{R}^{-1} \mathbf{G}_{t_i}^{(c)T} \mathbf{H}_{t_i}^{-1} \mathbf{G}_{t_i}^{(c)} \varepsilon_{t_i}.$$

■

The detailed version of the derivations on generalized path integral control as well as its iterative version can also be found in Theodorou (2011).

References

E. A. Theodorou. *Iterative Path Integral Stochastic Optimal Control: Theory and Applications to Motor Control.* PhD thesis, University of Southern California, May 2011.