A. Linear Time Hessian Inversion

In this supplementary appendix, we present an efficient algorithm for solving the linear system $\mathbf{h} = \mathbf{H}^{-1}\mathbf{g}$ in time linear in the path length T. As discussed in Section 5.1 of the paper, we solve the system with an upward solve of $(\tilde{\mathbf{H}}\mathbf{P} + \mathbf{J}\hat{\mathbf{H}})\bar{\mathbf{h}} = \mathbf{g}$ and a downward solve of $\mathbf{J}^{\mathrm{T}}\mathbf{h} = \bar{\mathbf{h}}$. The structure of $\mathbf{J}$ gives a downward pass that begins with $\mathbf{B}_1\mathbf{h}_1 = \bar{\mathbf{h}}_1$, followed by

$$\mathbf{B}_t\mathbf{h}_t = \bar{\mathbf{h}}_t - \mathbf{A}_t\bar{\mathbf{h}}_{t-1}. \tag{5}$$

Each row of the upward pass is given by

$$-\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t\bar{\mathbf{h}}_{t-1} + (\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger} + \mathbf{B}_t^{\mathrm{T}}\hat{\mathbf{H}}_t)\bar{\mathbf{h}}_t + \mathbf{B}_t^{\mathrm{T}}\mathbf{z}^{(t)} = \mathbf{g}_t,$$

where $\mathbf{z}^{(t)}$ is the sum over off-diagonal entries of $\mathbf{J}$, analogously to Equation 3. Since each row depends on both $\bar{\mathbf{h}}_t$ and $\bar{\mathbf{h}}_{t-1}$, we express $\bar{\mathbf{h}}_t$ as $\bar{\mathbf{h}}_t = \bar{\mathbf{h}}_t^{(1)} + \mathbf{S}_t\bar{\mathbf{h}}_{t-1}$, and seperately accumulate $\bar{\mathbf{h}}^{(1)}$ into $\mathbf{z}_{\mathbf{h}}$ and $\mathbf{S}$ into $\mathbf{z}_{\mathbf{S}}$, so that $\mathbf{z}^{(t)} = \mathbf{z}_{\mathbf{h}}^{(t)} + \mathbf{z}_{\mathbf{S}}^{(t)}\bar{\mathbf{h}}_t$. Collecting all $\bar{\mathbf{h}}_t$ terms on the left, we get the new row equation

$$(\underbrace{\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger} + \mathbf{B}_t^{\mathrm{T}}\hat{\mathbf{H}}_t + \mathbf{B}_t^{\mathrm{T}}\mathbf{z}_{\mathbf{S}}^{(t)}}_{\mathbf{D}_t})\bar{\mathbf{h}}_t = \mathbf{g}_t - \mathbf{B}_t^{\mathrm{T}}\mathbf{z}_{\mathbf{h}}^{(t)} + \tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t\bar{\mathbf{h}}_{t-1}.$$

When $\mathbf{x}$ and $\mathbf{u}$ have different dimensionality, $\mathbf{D}_t$ is not square, so the space of solutions for $\bar{\mathbf{h}}_t$ is given by

$$\bar{\mathbf{h}}_t = \underbrace{\mathbf{D}_t^{\dagger}(\mathbf{g}_t - \mathbf{B}_t^{\mathrm{T}}\mathbf{z}_{\mathbf{h}}^{(t)})}_{\bar{\mathbf{h}}_t^{(0)}} + \mathbf{D}_t^{\dagger}\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t\bar{\mathbf{h}}_{t-1} + \mathbf{N}_t\mathbf{u}_t, \tag{6}$$

where $\mathbf{N}_t$ is the null space of $\mathbf{D}_t$. Since $\bar{\mathbf{h}}_t$ must satisfy Equation 5, we can solve for $\mathbf{u}_t$ in terms of $\bar{\mathbf{h}}_{t-1}$ by substituting Equation 6 into Equation 5:

$$[\mathbf{B}_t, -\mathbf{N}_t][\mathbf{h}_t, \mathbf{u}_t]^{\mathrm{T}} = \bar{\mathbf{h}}_t^{(0)} + (\mathbf{D}_t^{\dagger}\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t - \mathbf{A}_t)\bar{\mathbf{h}}_{t-1}$$

The lower rows of this square system describe $\mathbf{u}_t$. Expressing the solution as $\mathbf{d}_t + \mathbf{C}_t\bar{\mathbf{h}}_{t-1}$ and substituting the lower rows into Equation 6, we obtain $\bar{\mathbf{h}}_t^{(1)}$ and $\mathbf{S}_t$:

$$\begin{aligned}\bar{\mathbf{h}}_t^{(1)} &= \bar{\mathbf{h}}_t^{(0)} + \mathbf{N}[\mathbf{d}_t]_{d_{\mathbf{u}}+1,\ldots,d_{\mathbf{x}}}\\ \mathbf{S}_t &= \mathbf{D}_t^{\dagger}\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t + \mathbf{N}_t[\mathbf{C}_t]_{d_{\mathbf{u}}+1,\ldots,d_{\mathbf{x}}}\end{aligned}$$

Finally, we update the accumulation variables $\mathbf{z}_{\mathbf{h}}$ and $\mathbf{z}_{\mathbf{S}}$ for the next step, analogously to Equation 3:

$$\begin{aligned}\mathbf{z}_{\mathbf{h}}^{(t-1)} &\leftarrow \mathbf{A}_t^{\mathrm{T}}((\hat{\mathbf{H}}_t + \mathbf{z}_{\mathbf{S}}^{(t)})\bar{\mathbf{h}}_t^{(1)} + \mathbf{z}_{\mathbf{h}}^{(t)})\\ \mathbf{z}_{\mathbf{S}}^{(t-1)} &\leftarrow \mathbf{A}_t^{\mathrm{T}}(\hat{\mathbf{H}}_t + \mathbf{z}_{\mathbf{S}}^{(t)})\mathbf{S}_t\end{aligned}$$

The log determinant $\log|-\mathbf{H}|$ is found by adding the log determinants of each row, given by $\log|-\mathbf{D}_t\mathbf{B}_t|$. The diagonal blocks $[\mathbf{J}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{J}]_t$ and $[\mathbf{H}^{-1}]_t$, which are used in Equation 4, can be computed during the downward pass. We first compute the following quantities:

$$\begin{aligned}\mathbf{Q}_t &= (\mathbf{I} + \mathbf{N}_t\,[\mathbf{B}_t, -\mathbf{N}_t]^{-1}_{d_{\mathbf{u}}+1,\ldots,d_{\mathbf{x}}})\mathbf{D}_t^{\dagger}\\ \mathbf{R}_t &= -(\mathbf{I} + \mathbf{N}_t\,[\mathbf{B}_t, -\mathbf{N}_t]^{-1}_{d_{\mathbf{u}}+1,\ldots,d_{\mathbf{x}}})\mathbf{D}_t^{\dagger}\mathbf{B}_t^{\mathrm{T}}\\ \mathbf{W}_t &= \mathbf{A}_t^{\mathrm{T}}(\hat{\mathbf{H}}_t + \mathbf{z}_{\mathbf{S}}^{(t)}),\end{aligned}$$

so that the upward pass for a new input matrix $\mathbf{b}$ is:

$$\bar{\mathbf{c}}_t^{(1)} = \mathbf{Q}_t\mathbf{b}_t + \mathbf{R}_t\mathbf{z}_{\mathbf{h}}^{(t)} \qquad \mathbf{z}_{\mathbf{h}}^{(t-1)} \leftarrow \mathbf{W}_t\bar{\mathbf{c}}_t^{(1)} + \mathbf{A}_t^{\mathrm{T}}\mathbf{z}_{\mathbf{h}}^{(t)}.$$

To evaluate the blocks of $\mathbf{H}^{-1}$ and $\mathbf{H}^{-1}\mathbf{J}$, we denote column t of the input (either $\mathbf{I}$ or $\mathbf{J}$) as $\mathbf{b}^t$, and the corresponding solution block $\mathbf{c}_t^t$. Since all blocks of $\mathbf{b}^t$ below t are zero, we can express $\mathbf{c}_t^t$ as a function of $\mathbf{b}_t^t$:

$$\mathbf{c}_t^t = \mathbf{B}_t^{\dagger}(\mathbf{Q}_t\mathbf{b}_t^t + (\mathbf{S}_t - \mathbf{A}_t)(\bar{\mathbf{c}}_{t-1}^t + \mathbf{V}_{t-1}\mathbf{W}_t\mathbf{Q}_t\mathbf{b}_t^t))$$

where $\bar{\mathbf{c}}_{t-1}^t$ is the intermediate solution if we were to set $\mathbf{b}_t^t$ to zero, and $\mathbf{V}_{t-1}$ is updated according to

$$\mathbf{V}_1 = \mathbf{R}_1 \qquad \mathbf{V}_t = \mathbf{R} + \mathbf{S}_t\mathbf{V}_{t-1}\mathbf{U}_t,$$

where $\mathbf{U}_t = \mathbf{W}_t\mathbf{R}_t + \mathbf{A}_t^{\mathrm{T}}$. For the blocks of $\mathbf{H}^{-1}$, $\bar{\mathbf{c}}_{t-1}^t$ is zero. These blocks are therefore given by

$$[\mathbf{H}^{-1}]_t = \mathbf{B}_t^{\dagger}(\mathbf{Q}_t + (\mathbf{S}_t - \mathbf{A}_t)\mathbf{V}_{t-1}\mathbf{W}_t\mathbf{Q}_t).$$

In the case of $\mathbf{H}^{-1}\mathbf{J}$, $\mathbf{b}_t^t = \mathbf{B}_t^{\mathrm{T}}$ and we have

$$\begin{aligned}\bar{\mathbf{c}}_{t-1}^t &= \bar{\mathbf{c}}_{t-1}^{t-1}\mathbf{A}_t^{\mathrm{T}} + \mathbf{V}_{t-1}\mathbf{W}_t\mathbf{Q}_t\mathbf{B}_t^{\mathrm{T}}\\ \bar{\mathbf{c}}_t^t &= \mathbf{Q}_t\mathbf{B}_t^{\mathrm{T}} + \mathbf{S}_t\bar{\mathbf{c}}_{t-1}^t.\end{aligned}$$

Algorithm 1 Linear Time Solution to $\mathbf{Hh} = \mathbf{g}$

Initialize $\mathbf{z}_{\mathbf{h}}^{(T)} \leftarrow 0$, $\mathbf{z}_{\mathbf{S}}^{(T)} \leftarrow 0$, $\det \leftarrow 0$
for $t = T$ **to** 1 **do**
  $\mathbf{D}_t = \mathbf{B}_t^{\mathrm{T}}\hat{\mathbf{H}}_t + \tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger} + \mathbf{B}_t^{\mathrm{T}}\mathbf{z}_{\mathbf{S}}^{(t)}$
  $\det \leftarrow \det + \log|-\mathbf{D}_t\mathbf{B}_t|$
  $\bar{\mathbf{h}}_t^{(0)} = \mathbf{D}_t^{\dagger}(\mathbf{g}_t - \mathbf{B}_t^{\mathrm{T}}\mathbf{z}_{\mathbf{h}}^{(t)})$
  $\mathbf{d}_t = [\mathbf{B}_t, -\mathbf{N}_t]^{-1}\,\bar{\mathbf{h}}_t^{(0)}$
  $\bar{\mathbf{h}}_t^{(1)} = \bar{\mathbf{h}}_t^{(0)} + \mathbf{N}_t[\mathbf{d}_t]_{d_{\mathbf{u}}+1,\ldots,d_{\mathbf{x}}}$
  if $t \neq 1$ **then**
    $\mathbf{C}_t = [\mathbf{B}_t, -\mathbf{N}_t]^{-1}\,(\mathbf{D}_t^{\dagger}\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t - \mathbf{A}_t)$
    $\mathbf{S}_t = \mathbf{D}_t^{\dagger}\tilde{\mathbf{H}}_t\mathbf{B}_t^{\dagger}\mathbf{A}_t + \mathbf{N}_t[\mathbf{C}_t]_{d_{\mathbf{u}}+1,\ldots,d_{\mathbf{x}}}$
    $\mathbf{z}_{\mathbf{h}}^{(t-1)} \leftarrow \mathbf{A}_t^{\mathrm{T}}\left((\hat{\mathbf{H}}_t + \mathbf{z}_{\mathbf{S}}^{(t)})\bar{\mathbf{h}}_t^{(1)} + \mathbf{z}_{\mathbf{h}}^{(t)}\right)$
    $\mathbf{z}_{\mathbf{S}}^{(t-1)} \leftarrow \mathbf{A}_t^{\mathrm{T}}\left(\hat{\mathbf{H}}_t + \mathbf{z}_{\mathbf{S}}^{(t)}\right)\mathbf{S}_t$
  else
    $\mathbf{h}_1 = [\mathbf{d}_1]_{1,\ldots,d_{\mathbf{u}}}$
    $\bar{\mathbf{h}}_1 = \bar{\mathbf{h}}_1^{(1)}$
  end if
end for
for $t = 2$ **to** T **do**
  $\bar{\mathbf{h}}_t = \bar{\mathbf{h}}_t^{(1)} + \mathbf{S}_t\bar{\mathbf{h}}_{t-1}$
  $\mathbf{h}_t = [\mathbf{d}_t]_{1,\ldots,d_{\mathbf{u}}} + [\mathbf{C}_t]_{1,\ldots,d_{\mathbf{u}}}\bar{\mathbf{h}}_{t-1}$
end for

Since we need the blocks $[\mathbf{H}_\mathbf{J}^{-1}]_t = [\mathbf{J}^\mathrm{T}\mathbf{H}^{-1}\mathbf{J}]_t$, we must also include the product of off-diagonal entries with the rows of $\mathbf{J}^\mathrm{T}$, using an auxiliary variable $\mathbf{F}_t$ to describe how this sum depends on $\mathbf{b}_t^t = \mathbf{B}_t^\mathrm{T}$:

$$[\mathbf{H}_\mathbf{J}^{-1}]_t = \mathbf{A}_t([\mathbf{H}_\mathbf{J}^{-1}]_{t-1}\mathbf{A}_t^\mathrm{T} + \mathbf{F}_{t-1}\mathbf{W}_t\mathbf{Q}_t\mathbf{B}_t^\mathrm{T}) + \mathbf{B}_t\mathbf{c}_t^t.$$

The auxiliary variable is updated to express the next sum as a function of $\mathbf{b}_{t+1}^{t+1}$ using $\mathbf{V}_{t-1}$ and $\mathbf{V}_t$:

$$\begin{aligned}\mathbf{F}_1 &= \mathbf{B}_1\mathbf{B}_1^\dagger\mathbf{V}_1 \\ \mathbf{F}_t &= \mathbf{A}_t\mathbf{F}_{t-1}\mathbf{U}_t + \mathbf{B}_t\mathbf{B}_t^\dagger(\mathbf{V}_t - \mathbf{A}_t\mathbf{V}_{t-1}\mathbf{U}_t).\end{aligned}$$

B. LQR Likelihood Derivation

In this section, we derive the MaxEnt likelihood for a quadratic approximation of the reward and linearized dynamics, which corresponds to the inverse LQR task. This likelihood can be viewed as an extension of the LQR IOC method described by Ziebart to arbitrary parameterized reward functions (Ziebart, 2010). We begin by redefining the states and actions in terms of their deviation from the example trajectory:

$$\bar{\mathbf{x}}_t = \mathbf{x}_t - \mathbf{x}_t^\star \qquad \bar{\mathbf{u}}_t = \mathbf{u}_t - \mathbf{u}_t^\star,$$

where $\mathbf{x}_t^\star$ and $\mathbf{u}_t^\star$ are the actions and states of the example path. The linearized dynamics are given by

$$\bar{\mathbf{x}}_t \approx \mathbf{A}_t\bar{\mathbf{x}}_{t-1} + \mathbf{B}_t\bar{\mathbf{u}}_t,$$

and the quadratic reward is given by

$$r(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) \approx r_t + \frac{1}{2}\bar{\mathbf{u}}_t^\mathrm{T}\tilde{\mathbf{H}}_t\bar{\mathbf{u}}_t + \bar{\mathbf{u}}_t^\mathrm{T}\tilde{\mathbf{g}}_t + \frac{1}{2}\bar{\mathbf{x}}_t\hat{\mathbf{H}}_t\bar{\mathbf{x}}_t + \bar{\mathbf{x}}_t^\mathrm{T}\hat{\mathbf{g}}_t.$$

Under the MaxEnt model, the probability of taking action $\bar{\mathbf{u}}_t$ in state $\bar{\mathbf{x}}_{t-1}$ is given by

$$P(\bar{\mathbf{u}}_t|\bar{\mathbf{x}}_{t-1}) = \exp\left(Q_t(\bar{\mathbf{x}}_{t-1}, \bar{\mathbf{u}}_t) - V_t(\bar{\mathbf{x}}_{t-1})\right),$$

where $Q_t(\bar{\mathbf{x}}_{t-1}, \bar{\mathbf{u}}_t) = r(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) + V_{t+1}(\bar{\mathbf{x}}_t)$ and the value function V_t is defined with a "softened" form of the Bellman backup (Ziebart, 2010):

$$V_t(\bar{\mathbf{x}}_{t-1}) = \log\int e^{Q_t(\bar{\mathbf{x}}_{t-1}, \bar{\mathbf{u}}_t)}d\bar{\mathbf{u}}_t.$$

In LQR, we assume that we can express $V_t(\bar{\mathbf{x}}_{t-1})$ as a quadratic form:

$$V_t(\bar{\mathbf{x}}_{t-1}) = \frac{1}{2}\bar{\mathbf{x}}_{t-1}^\mathrm{T}\hat{\mathbf{V}}_t\bar{\mathbf{x}}_{t-1} + \bar{\mathbf{x}}_{t-1}^\mathrm{T}\hat{\mathbf{v}}_t.$$

Substituting in this equation for V_t, the quadratic version of $r(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)$, and the linearized dynamics, we get

$$\begin{aligned}Q_t(\bar{\mathbf{x}}_{t-1}, \bar{\mathbf{u}}_t) = r_t &+ \frac{1}{2}\bar{\mathbf{u}}_t^\mathrm{T}\tilde{\mathbf{H}}_t\bar{\mathbf{u}}_t + \bar{\mathbf{u}}_t^\mathrm{T}\tilde{\mathbf{g}}_t \\ &+ \frac{1}{2}\left[\mathbf{A}_t\bar{\mathbf{x}}_{t-1} + \mathbf{B}_t\bar{\mathbf{u}}_t\right]^\mathrm{T}\underbrace{\left[\hat{\mathbf{H}}_t + \hat{\mathbf{V}}_t\right]}_{\hat{\mathbf{Q}}_t}\left[\mathbf{A}_t\bar{\mathbf{x}}_{t-1} + \mathbf{B}_t\bar{\mathbf{u}}_t\right] \\ &+ \left[\mathbf{A}_t\bar{\mathbf{x}}_{t-1} + \mathbf{B}_t\bar{\mathbf{u}}_t\right]^\mathrm{T}\underbrace{\left[\hat{\mathbf{g}}_t + \hat{\mathbf{v}}_t\right]}_{\hat{\mathbf{q}}_t}.\end{aligned}$$

To determine for $V_t(\bar{\mathbf{x}}_{t-1})$, we solve the integral analytically by rewriting it as an unnormalized Gaussian. This produces the following equation:

$$\begin{aligned}V_t(\bar{\mathbf{x}}_{t-1}) = r_t &+ \frac{1}{2}\bar{\mathbf{x}}_{t-1}^\mathrm{T}\mathbf{A}_t^\mathrm{T}\hat{\mathbf{Q}}_t\mathbf{A}_t\bar{\mathbf{x}}_{t-1} + \bar{\mathbf{x}}_{t-1}^\mathrm{T}\mathbf{A}_t^\mathrm{T}\hat{\mathbf{q}}_t \\ &- \frac{1}{2}\tilde{\mathbf{m}}_t^\mathrm{T}\tilde{\mathbf{M}}_t^{-1}\tilde{\mathbf{m}}_t - \frac{1}{2}\log\left|-\tilde{\mathbf{M}}_t\right|,\end{aligned}$$

where we define $\tilde{\mathbf{m}}_t = \tilde{\mathbf{g}}_t + \mathbf{B}_t^\mathrm{T}\hat{\mathbf{q}}_t + \mathbf{B}_t^\mathrm{T}\hat{\mathbf{Q}}_t\mathbf{A}_t\bar{\mathbf{x}}_{t-1}$ and $\tilde{\mathbf{M}}_t = \tilde{\mathbf{H}}_t + \mathbf{B}_t^\mathrm{T}\hat{\mathbf{Q}}_t\mathbf{B}_t$. To get the likelihood of the observed action, we simply subtract the value function from the Q function, resulting in

$$\log P(\mathbf{u}_t|\mathbf{x}_{t-1}) = \mathcal{L}_t = \frac{1}{2}\tilde{\mathbf{m}}_t^\mathrm{T}\tilde{\mathbf{M}}_t^{-1}\tilde{\mathbf{m}}_t + \frac{1}{2}\log\left|-\tilde{\mathbf{M}}_t\right|.$$

We can also rewrite the equation for V_t as a quadratic form in $\bar{\mathbf{x}}_{t-1}$ to get the value function matrix and vector for the preceding time step:

$$\begin{aligned}\hat{\mathbf{V}}_t =&\mathbf{A}_{t+1}^\mathrm{T}\hat{\mathbf{Q}}_{t+1}\mathbf{A}_{t+1} - \\ &\mathbf{A}_{t+1}^\mathrm{T}\hat{\mathbf{Q}}_{t+1}^\mathrm{T}\mathbf{B}_{t+1}\tilde{\mathbf{M}}_{t+1}^{-1}\mathbf{B}_{t+1}^\mathrm{T}\hat{\mathbf{Q}}_{t+1}\mathbf{A}_{t+1} \\ \hat{\mathbf{v}}_t =&\mathbf{A}_{t+1}^\mathrm{T}\hat{\mathbf{q}}_{t+1} - \\ &\mathbf{A}_{t+1}^\mathrm{T}\hat{\mathbf{Q}}_{t+1}\mathbf{B}_{t+1}\tilde{\mathbf{M}}_{t+1}^{-1}\left[\tilde{\mathbf{g}}_{t+1} + \mathbf{B}_{t+1}^\mathrm{T}\hat{\mathbf{q}}_{t+1}\right].\end{aligned}$$

The complete likelihood is then given by

$$\mathcal{L} = \sum_t \frac{1}{2}\tilde{\mathbf{m}}_t^\mathrm{T}\tilde{\mathbf{M}}_t^{-1}\tilde{\mathbf{m}}_t + \frac{1}{2}\log\left|-\tilde{\mathbf{M}}_t\right|,$$

and the corresponding gradient is

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\theta_i} = \sum_t &\tilde{\mathbf{m}}_t^\mathrm{T}\tilde{\mathbf{M}}_t^{-1}\frac{\partial\tilde{\mathbf{m}}_t}{\partial\theta_i} + \frac{1}{2}\mathrm{tr}\left(\tilde{\mathbf{M}}_t^{-1}\frac{\partial\tilde{\mathbf{M}}_t}{\partial\theta_i}\right) \\ &- \frac{1}{2}\tilde{\mathbf{m}}_t^\mathrm{T}\tilde{\mathbf{M}}_t^{-1}\frac{\partial\tilde{\mathbf{M}}_t}{\partial\theta_i}\tilde{\mathbf{M}}_t^{-1}\tilde{\mathbf{m}}_t.\end{aligned}$$

The derivatives of $\tilde{\mathbf{M}}_t$ and $\tilde{\mathbf{m}}_t$ are

$$\frac{\partial\tilde{\mathbf{m}}_t}{\partial\theta_i} = \frac{\partial\tilde{\mathbf{g}}_t}{\partial\theta_i} + \mathbf{B}_t^\mathrm{T}\frac{\partial\hat{\mathbf{q}}_t}{\partial\theta_i} \qquad \frac{\partial\tilde{\mathbf{M}}_t}{\partial\theta_i} = \frac{\partial\tilde{\mathbf{H}}_t}{\partial\theta_i} + \mathbf{B}_t^\mathrm{T}\frac{\partial\hat{\mathbf{Q}}_t}{\partial\theta_i}\mathbf{B}_t,$$

and the derivatives of $\hat{\mathbf{Q}}_t$ and $\hat{\mathbf{q}}_t$ are given recursively:

$$\frac{\partial \hat{\mathbf{q}}_t}{\partial \theta_i} = \frac{\partial \hat{\mathbf{g}}_t}{\partial \theta_i} + \mathbf{A}_{t+1}^{\mathrm{T}} \frac{\partial \hat{\mathbf{q}}_{t+1}}{\partial \theta_i}$$
$$- \mathbf{A}_{t+1}^{\mathrm{T}} \hat{\mathbf{Q}}_{t+1} \mathbf{B}_{t+1} \tilde{\mathbf{M}}_{t+1}^{-1} \left[\frac{\partial \tilde{\mathbf{g}}_{t+1}}{\partial \theta_i} + \mathbf{B}_{t+1}^{\mathrm{T}} \frac{\partial \hat{\mathbf{q}}_{t+1}}{\partial \theta_i} \right]$$
$$- \mathbf{A}_{t+1}^{\mathrm{T}} \frac{\partial \hat{\mathbf{Q}}_{t+1}}{\partial \theta_i} \mathbf{B}_{t+1} \tilde{\mathbf{M}}_{t+1}^{-1} \left[\tilde{\mathbf{g}}_{t+1} + \mathbf{B}_{t+1}^{\mathrm{T}} \hat{\mathbf{q}}_{t+1} \right]$$
$$+ \mathbf{A}_{t+1}^{\mathrm{T}} \hat{\mathbf{Q}}_{t+1} \mathbf{B}_{t+1} \tilde{\mathbf{M}}_{t+1}^{-1} \frac{\partial \tilde{\mathbf{M}}_{t+1}}{\partial \theta_i} \tilde{\mathbf{M}}_{t+1}^{-1} \left[\tilde{\mathbf{g}}_{t+1} + \mathbf{B}_{t+1}^{\mathrm{T}} \hat{\mathbf{q}}_{t+1} \right]$$
$$\frac{\partial \hat{\mathbf{Q}}_t}{\partial \theta_i} = \frac{\partial \hat{\mathbf{H}}_t}{\partial \theta_i} + \mathbf{A}_{t+1}^{\mathrm{T}} \frac{\partial \hat{\mathbf{Q}}_{t+1}}{\partial \theta_i} \mathbf{A}_{t+1}$$
$$- \mathbf{A}_{t+1}^{\mathrm{T}} \hat{\mathbf{Q}}_{t+1}^{\mathrm{T}} \mathbf{B}_{t+1} \tilde{\mathbf{M}}_{t+1}^{-1} \mathbf{B}_{t+1}^{\mathrm{T}} \frac{\partial \hat{\mathbf{Q}}_{t+1}}{\partial \theta_i} \mathbf{A}_{t+1}$$
$$- \mathbf{A}_{t+1}^{\mathrm{T}} \frac{\partial \hat{\mathbf{Q}}_{t+1}}{\partial \theta_i}^{\mathrm{T}} \mathbf{B}_{t+1} \tilde{\mathbf{M}}_{t+1}^{-1} \mathbf{B}_{t+1}^{\mathrm{T}} \hat{\mathbf{Q}}_{t+1} \mathbf{A}_{t+1}$$
$$+ \mathbf{A}_{t+1}^{\mathrm{T}} \hat{\mathbf{Q}}_{t+1}^{\mathrm{T}} \mathbf{B}_{t+1} \tilde{\mathbf{M}}_{t+1}^{-1} \frac{\partial \tilde{\mathbf{M}}_{t+1}}{\partial \theta_i} \tilde{\mathbf{M}}_{t+1}^{-1} \mathbf{B}_{t+1}^{\mathrm{T}} \hat{\mathbf{Q}}_{t+1} \mathbf{A}_{t+1}.$$

Experimentally, we confirmed that this algorithm produces the same likelihood and gradients as the linear time Hessian inversion algorithm. It lacks the convenient close-form gradient formula in Equation 4, which can make gradient evaluation more tricky in settings with highly structured gradients, such as the Gaussian process gradients discussed in the following section. However, the recursion itself is somewhat simpler to implement than the direct Hessian inversion.

C. Nonlinear Likelihood Gradients

In this appendix, we present an efficient method for computing the gradients of the likelihood for the nonlinear variant of our method. We first introduce the following intermediate quantities:

$$\Delta_{tik} = \mathbf{f}_k^i - \mathbf{f}_k^t \quad \text{and} \quad \hat{\xi}_{tip} = \sum_{k,i} \hat{\mathbf{g}}^{(k)} \lambda_k \Delta_{tik}.$$

Using these quantities, the gradient $\hat{\mathbf{g}}$ and Hessian $\hat{\mathbf{H}}$ are given by

$$\hat{\mathbf{g}}_{tp} = \sum_{k,i} \hat{\mathbf{g}}_{tp}^{(k)} \lambda_k \Delta_{tik} \mathbf{k}_{ti} \alpha_i$$
$$\hat{\mathbf{H}}_{tpq} = \sum_i \left(\hat{\xi}_{tip} \hat{\xi}_{tiq} + \sum_k \lambda_k (\hat{\mathbf{H}}_{tpq}^{(k)} \Delta_{tik} + \hat{\mathbf{g}}_{tip}^{(k)} \hat{\mathbf{g}}_{tiq}^{(k)}) \right) \mathbf{k}_{ti} \alpha_i.$$

The terms $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{H}}$ are given analogously. The kernel variance β factors out of these equations, so the likelihood gradient with respect to β depends only on the GP prior. For $\mathbf{y}$ and λ, we could simply differentiate $\hat{\mathbf{g}}$ and $\hat{\mathbf{H}}$ and apply Equation 4. However, we can compute the gradient more efficiently by noting that only the feature derivatives $\hat{\mathbf{g}}^{(k)}$ and $\hat{\mathbf{H}}^{(k)}$ depend on the state dimensions. This allows us to compute the likelihood gradients without computing the gradients of $\hat{\mathbf{g}}$ and $\hat{\mathbf{H}}$. Let $\Gamma_{tp}^{\hat{\mathbf{g}}}$ and $\Gamma_{tpq}^{\hat{\mathbf{H}}}$ denote the coefficients of $\hat{\mathbf{g}}_{tp}$ and $\hat{\mathbf{H}}_{tpq}$ in Equation 4. We can then rewrite it as

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{tp} \frac{\partial \tilde{\mathbf{g}}_{tp}}{\partial \theta} \Gamma_{tp}^{\tilde{\mathbf{g}}} + \sum_{tp} \frac{\partial \hat{\mathbf{g}}_{tp}}{\partial \theta} \Gamma_{tp}^{\hat{\mathbf{g}}} + \sum_{tpq} \frac{\partial \tilde{\mathbf{H}}_{tij}}{\partial \theta} \Gamma_{tpq}^{\tilde{\mathbf{H}}} + \sum_{tpq} \frac{\partial \hat{\mathbf{H}}_{tij}}{\partial \theta} \Gamma_{tpq}^{\hat{\mathbf{H}}} + \sum_{tp} \frac{\partial \hat{\mathbf{g}}_{ti}}{\partial \theta} \Gamma_{tp}^{\check{\mathbf{H}}}.$$

Now we precompute the following quantities:

$$\nu_{tk} = \sum_p \tilde{\mathbf{g}}_{tp}^{(k)} \Gamma_{tp}^{\tilde{\mathbf{g}}} + \sum_p \hat{\mathbf{g}}_{tp}^{(k)} \Gamma_{tp}^{\hat{\mathbf{g}}}$$
$$\rho_{tk} = \sum_{pq} \tilde{\mathbf{g}}_{tp}^{(k)} \tilde{\mathbf{g}}_{tq}^{(k)} \Gamma_{tpq}^{\tilde{\mathbf{H}}} + \sum_{pq} \hat{\mathbf{g}}_{tp}^{(k)} \hat{\mathbf{g}}_{tq}^{(k)} \Gamma_{tpq}^{\hat{\mathbf{H}}}$$
$$\omega_{tk} = \sum_{pq} \check{\mathbf{H}}_{tpq}^{(k)} \Gamma_{tpq}^{\tilde{\mathbf{H}}} + \sum_{pq} \hat{\mathbf{H}}_{tpq}^{(k)} \Gamma_{tpq}^{\hat{\mathbf{H}}} + \sum_p \hat{\mathbf{g}}_{tp}^{(k)} \Gamma_{tp}^{\check{\mathbf{H}}}.$$

For the gradient with respect to $\mathbf{y}$, we first compute

$$\zeta_{ti}^{(\mathbf{y})} = \sum_{pq} \Gamma_{tpq}^{\hat{\mathbf{H}}} \hat{\xi}_{tip} \hat{\xi}_{tiq} + \sum_{pq} \Gamma_{tpq}^{\tilde{\mathbf{H}}} \tilde{\xi}_{tip} \tilde{\xi}_{tiq} + \sum_k \lambda_k ((\nu_{tk} + \omega_{tk}) \Delta_{tik} - \rho_{tk}).$$

The gradient with respect to $\mathbf{y}_j$ is now given by

$$\frac{\partial \mathcal{L}}{\partial \mathbf{y}_j} = \sum_{ti} \zeta_{ti}^{(\mathbf{y})} \mathbf{k}_{ti} [\mathbf{K}^{-1}]_{ij} + \frac{\partial}{\partial \mathbf{y}_j} \log P(\mathbf{y}, \lambda, \beta | \mathbf{F}).$$

where the GP prior gradient is found in previous work (Levine et al., 2011). For the gradient with respect to λ_k, we compute another auxiliary quantity:

$$\zeta_{tik}^{(\lambda)} = \nu_{tk} + \omega_{tk} + 2 \sum_{pq} \Gamma_{tpq}^{\hat{\mathbf{H}}} \hat{\xi}_{tip} \hat{\mathbf{g}}_{tq}^{(k)} + 2 \sum_{pq} \Gamma_{tpq}^{\tilde{\mathbf{H}}} \tilde{\xi}_{tip} \tilde{\mathbf{g}}_{tq}^{(k)}$$

The gradient with respect to λ_k is then given by

$$\frac{\partial \mathcal{L}}{\partial \lambda_k} = \left[\sum_{ti} \left(\zeta_{tik}^{(\lambda)} \Delta_{tik} - \frac{1}{2} \zeta_{ti}^{(\mathbf{y})} \Delta_{tik}^{(2)} - \rho_{tk} \right) \mathbf{k}_{ti} \alpha_i \right] + \left[\sum_{ti} \zeta_{ti}^{(\mathbf{y})} \mathbf{k}_{ti} \frac{\partial \alpha_i}{\partial \lambda_k} \right] + \frac{\partial}{\partial \lambda_k} \log P(\mathbf{y}, \lambda, \beta | \mathbf{F}),$$

where $\Delta_{tik}^{(2)} = (\mathbf{f}_k^i - \mathbf{f}_k^t)^2 + \sigma^2$.

The advantage of this approach is that we avoid forming any factor with size dependent on both the dimensionality of the state space and the number of inducing points, as well as any factor that depends quadratically on the number of features. This allows the algorithm to scale gracefully with state and action dimensionality, inducing point count, and feature count.