Variational Inference for Inverse Reinforcement Learning with Gaussian Processes: Supplementary Material

Paulius Dilkas (2146879)

6th December 2018

1 Proofs

We primarily think of rewards as a vector $\mathbf{r} \in \mathbb{R}^{|\mathcal{S}|}$, but sometimes we use a function notation $r(s)$ to denote the reward of a particular state $s \in \mathcal{S}$. The functional notation is purely a notational convenience.

MDP values are characterised by both a state and a reward function/vector. In order to prove the next theorem, we think of the value function as $V : \mathcal{S} \to \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}$, i.e., V takes a state $s \in \mathcal{S}$ and returns a function $V(s) : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}$ that takes a reward vector $\mathbf{r} \in \mathbb{R}^{|\mathcal{S}|}$ and returns a value of the state s, $V_{\mathbf{r}}(s) \in \mathbb{R}$. The function $V(s)$ computes the values of all states and returns the value of state s.

Theorem 1. *MDP value functions $V(s) : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}$ (for $s \in \mathcal{S}$) are Lebesgue measurable.*

Proof sketch. For any reward vector $\mathbf{r} \in \mathbb{R}^{|\mathcal{S}|}$, the collection of converged value functions $\{V_{\mathbf{r}}(s) \mid s \in \mathcal{S}\}$ satisfy

$$\forall s \in \mathcal{S},\ V_{\mathbf{r}}(s) = \log \sum_{a \in \mathcal{A}} \exp \left(r(s) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s, a, s') V_{\mathbf{r}}(s') \right).$$

Let $s_i \in \mathcal{S}$ be an arbitrary state. In order to prove that $V(s_i)$ is measurable, it is enough to show that for any $\alpha \in \mathbb{R}$, the set

$$\Bigg\{ \mathbf{r} \in \mathbb{R}^{|\mathcal{S}|} \mid V_{\mathbf{r}}(s_i) \in (-\infty, \alpha);$$
$$\forall s \in \mathcal{S} \setminus \{s_i\},\ V_{\mathbf{r}}(s) \in \mathbb{R};$$
$$\forall s \in \mathcal{S},\ V_{\mathbf{r}}(s) = \log \sum_{a \in \mathcal{A}} \exp \left(r(s) + \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s, a, s') V_{\mathbf{r}}(s') \right) \Bigg\}$$

is measurable. Since this set can be constructed in Zermelo-Fraenkel set theory *without* the axiom of choice, it is measurable [1], which proves that $V(s)$ is a measurable function for any $s \in \mathcal{S}$. □

Definition 1. For any finite-dimensional vector $\mathbf{x} = (x_1, \dots, x_n)^\intercal$, its *maximum norm* is

$$\|\mathbf{x}\|_\infty = \max_i |x_i|.$$

Theorem 2. *If the initial values of the MDP value function satisfy the following bound, then the bound remains satisfied throughout value iteration:*

$$|V_{\mathbf{r}}(s)| \le \frac{\|\mathbf{r}\|_\infty + \log |\mathcal{A}|}{1 - \gamma}. \tag{1}$$

Proof. We begin by considering (1) without taking the absolute value of $V_{\mathbf{r}}(s)$, i.e.,

$$V_{\mathbf{r}}(s) \leq \frac{\|\mathbf{r}\|_\infty + \log|\mathcal{A}|}{1-\gamma}, \tag{2}$$

and assuming that the initial values of $\{V_{\mathbf{r}}(s) \mid s \in \mathcal{S}\}$ already satisfy (2). For each $s \in \mathcal{S}$, the value of $V_{\mathbf{r}}(s)$ is updated via this rule:

$$V_{\mathbf{r}}(s) \coloneqq \log \sum_{a\in\mathcal{A}} \exp\left(r(s) + \gamma \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') V_{\mathbf{r}}(s')\right).$$

Note that both log and exp are increasing functions, $\gamma > 0$, and the $\mathcal{T}$ function gives a probability (a non-negative number). Thus

$$\begin{aligned}
V_{\mathbf{r}}(s) &\leq \log \sum_{a\in\mathcal{A}} \exp\left(r(s) + \gamma \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s') \frac{\|\mathbf{r}\|_\infty + \log|\mathcal{A}|}{1-\gamma}\right) \\
&= \log \sum_{a\in\mathcal{A}} \exp\left(r(s) + \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|)}{1-\gamma} \sum_{s'\in\mathcal{S}} \mathcal{T}(s,a,s')\right) \\
&= \log \sum_{a\in\mathcal{A}} \exp\left(r(s) + \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|)}{1-\gamma}\right)
\end{aligned}$$

by the definition of $\mathcal{T}$. Then

$$\begin{aligned}
V_{\mathbf{r}}(s) &\leq \log\left(|\mathcal{A}| \exp\left(r(s) + \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|)}{1-\gamma}\right)\right) \\
&= \log\left(\exp\left(\log|\mathcal{A}| + r(s) + \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|)}{1-\gamma}\right)\right) \\
&= \log|\mathcal{A}| + r(s) + \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|)}{1-\gamma} \\
&= \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|) + (1-\gamma)(\log|\mathcal{A}| + r(s))}{1-\gamma} \\
&\leq \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|) + (1-\gamma)(\log|\mathcal{A}| + \|\mathbf{r}\|_\infty)}{1-\gamma} \\
&= \frac{\|\mathbf{r}\|_\infty + \log|\mathcal{A}|}{1-\gamma}
\end{aligned}$$

by the definition of $\|\mathbf{r}\|_\infty$.

The proof for

$$V_{\mathbf{r}}(s) \geq \frac{\|\mathbf{r}\|_\infty + \log|\mathcal{A}|}{\gamma-1} \tag{3}$$

follows the same argument until we get to

$$\begin{aligned}
V_{\mathbf{r}}(s) &\geq \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|) + (\gamma-1)(\log|\mathcal{A}| + r(s))}{\gamma-1} \\
&\geq \frac{\gamma(\|\mathbf{r}\|_\infty + \log|\mathcal{A}|) + (\gamma-1)(-\log|\mathcal{A}| - \|\mathbf{r}\|_\infty)}{\gamma-1} \\
&= \frac{\|\mathbf{r}\|_\infty + \log|\mathcal{A}|}{\gamma-1},
\end{aligned}$$

where we use the fact that $r(s) \geq -\|\mathbf{r}\|_\infty - 2\log|\mathcal{A}|$. Combining (2) and (3) gives (1). $\square$

Theorem 3 (The Lebesgue Dominated Convergence Theorem [3])**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{f_n\}$ a sequence of measurable functions on X for which $\{f_n\} \to f$ pointwise a.e. on X and the function f is measurable. Assume there is a non-negative function g that is integrable over X and dominates the sequence $\{f_n\}$ on X in the sense that*

$$|f_n| \leq g \text{ a.e. on } X \text{ for all } n.$$

Then f is integrable over X and

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Proposition 1 ([3])**.** *Let f be a measurable function on E. Suppose there is a non-negative function g that is integrable over E and dominates f in the sense that*

$$|f| \leq g \text{ on } E.$$

Then f is integrable over E.

Theorem 4. *Using our usual notation,*

$$\frac{\partial}{\partial t} \iiint V_{\mathbf{r}}(s) q(\mathbf{r}) q(\mathbf{u}) q(\boldsymbol{\lambda}) \, d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda} = \iiint \frac{\partial}{\partial t} [V_{\mathbf{r}}(s) q(\mathbf{r}) q(\mathbf{u}) q(\boldsymbol{\lambda})] \, d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda},$$

where $t \in \{\mathbf{m}, \mathbf{S}, \alpha_0, \ldots, \alpha_d, \beta_0, \ldots, \beta_d\}$.

Proof. Let

$$\begin{aligned} f(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t) &= V_{\mathbf{r}}(s) q(\mathbf{r}) q(\mathbf{u}) q(\boldsymbol{\lambda}), \\ F(t) &= \iiint f(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t) \, d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda}, \end{aligned}$$

and, for any t, let $(t_n)_{n=1}^{\infty}$ be any sequence such that $\lim_{n\to\infty} t_n = t$, but $t_n \neq t$ for all n. We want to show that

$$F'(t) = \lim_{n\to\infty} \frac{F(t_n) - F(t)}{t_n - t} = \iiint \left. \frac{\partial f}{\partial t} \right|_{(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t)} d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda}. \tag{4}$$

We have

$$\frac{F(t_n) - F(t)}{t_n - t} = \iiint \frac{f(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t_n) - f(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t)}{t_n - t} \, d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda} = \iiint f_n(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}) \, d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda},$$

where

$$f_n(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}) = \frac{f(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t_n) - f(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t)}{t_n - t}.$$

Since

$$\lim_{n\to\infty} f_n(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}) = \left. \frac{\partial f}{\partial t} \right|_{(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}, t)},$$

(4) follows from Theorem 3 as soon as we show that both f and f_n are measurable and find a non-negative integrable function g such that for all n, $\mathbf{r}$, $\mathbf{u}$, $\boldsymbol{\lambda}$,

$$|f_n(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda})| \leq g(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda}).$$

The MDP value function is measurable by Theorem 1. The result of multiplying or adding measurable functions (e.g., probability density functions (PDFs)) to a measurable function is still measurable. Thus, both f and f_n are measurable.

It remains to find g. Without loss of generality, assume that t is a parameter of $q(\boldsymbol{\lambda})$. Then

$$|f_n(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda})| = |V_{\mathbf{r}}(s) q(\mathbf{r}) q(\mathbf{u}) q(\boldsymbol{\lambda})| = |V_{\mathbf{r}}(s)| q(\mathbf{r}) q(\mathbf{u}) \left| \frac{q(\boldsymbol{\lambda})|_{t=t_n} - q(\boldsymbol{\lambda})}{t_n - t} \right|$$

since PDFs are non-negative. An upper bound for $|V_{\mathbf{r}}(s)|$ is given by Theorem 2, while

$$\frac{q(\boldsymbol{\lambda})|_{t=t_n} - q(\boldsymbol{\lambda})}{t_n - t} = \left. \frac{\partial q(\boldsymbol{\lambda})}{\partial t} \right|_{t=c}$$

for some c between t and t_n due to the mean value theorem (since q is a continuous and differentiable function of t, regardless of the specific choices of q and t). Therefore,

$$|f_n(\mathbf{r}, \mathbf{u}, \boldsymbol{\lambda})| \leq \frac{\|\mathbf{r}\|_\infty + \log |\mathcal{A}|}{1 - \gamma} q(\mathbf{r}) q(\mathbf{u}) \left| \left. \frac{\partial q(\boldsymbol{\lambda})}{\partial t} \right|_{t=c} \right|.$$

The bound is clearly non-negative and measurable. It remains to show that it is also integrable. Let $\mathbf{r} = (r_1, \ldots, r_k)^\intercal$. Then

$$\frac{\|\mathbf{r}\|_\infty + \log |\mathcal{A}|}{1 - \gamma} q(\mathbf{r}) \leq \frac{q(\mathbf{r})}{1 - \gamma} \left(\log |\mathcal{A}| + \sum_{i=1}^{k} |r_i| \right),$$

and

$$\int \frac{q(\mathbf{r})}{1 - \gamma} \left(\log |\mathcal{A}| + \sum_{i=1}^{k} |r_i| \right) d\mathbf{r} = \frac{\log |\mathcal{A}|}{1 - \gamma} + \frac{1}{1 - \gamma} \sum_{i=1}^{k} \mathbb{E}[|r_i|],$$

which clearly exists and is finite, so

$$\int \frac{\|\mathbf{r}\|_\infty + \log |\mathcal{A}|}{1 - \gamma} q(\mathbf{r}) \, d\mathbf{r} < \infty$$

for all γ and $\mathbf{u}$. The existence of

$$\iiint \frac{\|\mathbf{r}\|_\infty + \log |\mathcal{A}|}{1 - \gamma} q(\mathbf{r}) q(\mathbf{u}) \left| \left. \frac{\partial q(\boldsymbol{\lambda})}{\partial t} \right|_{t=c} \right| \, d\mathbf{r} \, d\mathbf{u} \, d\boldsymbol{\lambda}$$

then comes from a boundedness argument, as... □

Theorem 5 (Differentiating under the integral sign)**.** *Assume $f : R \times R \to R$ is such that $x \mapsto f(x,t)$ is measurable for each $t \in R$, that $f(x, t_0)$ is integrable for some $t_0 \in R$ and $\frac{\partial f(x,t)}{\partial t}$ exists for each (x, t). Assume also that there is an integrable $g : R \to R$ with $\left|\frac{\partial f(x,t)}{\partial t}\right| \leq g(x)$ for each $x, t \in R$. Then the function $x \mapsto f(x, t)$ is integrable for each t and the function $F : R \to R$ defined by*

$$F(t) = \int_R f_t \, d\mu = \int_R f(x,t) \, d\mu(x)$$

is differentiable with derivative

$$F'(t) = \frac{d}{dt} \int_R f(x,t) \, d\mu(x) = \int_R \frac{\partial}{\partial t} f(x,t) \, d\mu(x).$$

Proof. To prove the formula for $F(t)$ consider any sequence $(t_n)_{n=1}^{\infty}$ so that $\lim_{n\to\infty} t_n = t$ but $t_n \neq t$ for each t. We claim that

$$\lim_{n\to\infty} \frac{F(t_n) - F(t)}{t_n - t} = \int_R \frac{\partial f(x,t)}{\partial t} \, d\mu(x). \tag{5}$$

We have

$$\frac{F(t_n) - F(t)}{t_n - t} = \int_R \frac{f(x, t_n) - f(x, t)}{t_n - t}\, d\mu(x) = \int_R f_n(x)\, d\mu(x)$$

where

$$f_n(x) = \frac{f(x, t_n) - f(x, t)}{t_n - t}.$$

Notice that, for each x we know

$$\lim_{n\to\infty} f_n(x) = \left.\frac{\partial f}{\partial t}\right|_{(x,t)}$$

and so (5) will follow from the dominated convergence theorem once we show that $|f_n(x)| \leq g(x)$ for each x.

That follows from the mean value theorem again because there is c between t and t_0 (with c depending on x) so that

$$f_n(x) = \frac{f(x, t_n) - f(x, t)}{t_n - t} = \left.\frac{\partial f}{\partial t}\right|_{(x,c)}.$$

So $|f_n(x)| \leq g(x)$ for each x. □

2 Derivatives of the Evidence Lower Bound

2.1 $\partial/\partial\mathbf{m}$

We begin by removing terms independent of $\mathbf{m}$:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{m}} =& -\frac{1}{2}\frac{\partial}{\partial\mathbf{m}}[\mathbf{m}^\intercal\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]\mathbf{m}] + \frac{\partial}{\partial\mathbf{m}}[\mathbf{t}^\intercal\mathbb{E}[\mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]\mathbf{m}] \\ &- \sum_{i=1}^N\sum_{t=1}^T \frac{\partial}{\partial\mathbf{m}}\mathbb{E}[V_\mathbf{r}(s_{i,t})] - \gamma\sum_{s'\in\mathcal{S}}\mathcal{T}(s_{i,t}, a_{i,t}, s')\frac{\partial}{\partial\mathbf{m}}\mathbb{E}[V_\mathbf{r}(s')].\end{aligned}$$

Here

$$\begin{aligned}\frac{\partial}{\partial\mathbf{m}}[\mathbf{m}^\intercal\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]\mathbf{m}] &= (\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}] + \mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]^\intercal)\mathbf{m}, \\ \frac{\partial}{\partial\mathbf{m}}[\mathbf{t}^\intercal\mathbb{E}[\mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]\mathbf{m}] &= \mathbf{t}^\intercal\mathbb{E}[\mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}],\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial\mathbf{m}}\mathbb{E}[V_\mathbf{r}(s)] &= \frac{\partial}{\partial\mathbf{m}}\iiint V_\mathbf{r}(s)p(\mathbf{r}|\boldsymbol{\lambda}, \mathbf{X}_\mathbf{u}, \mathbf{u})\mathcal{N}(\mathbf{u}; \mathbf{m}, \mathbf{S})q(\boldsymbol{\lambda})\, d\mathbf{r}\, d\mathbf{u}\, d\boldsymbol{\lambda} \\ &= \iiint V_\mathbf{r}(s)p(\mathbf{r}|\boldsymbol{\lambda}, \mathbf{X}_\mathbf{u}, \mathbf{u})\frac{\partial}{\partial\mathbf{m}}\mathcal{N}(\mathbf{u}; \mathbf{m}, \mathbf{S})q(\boldsymbol{\lambda})\, d\mathbf{r}\, d\mathbf{u}\, d\boldsymbol{\lambda},\end{aligned} \tag{6}$$

where

$$\begin{aligned}\frac{\partial}{\partial\mathbf{m}}\mathcal{N}(\mathbf{u}; \mathbf{m}, \mathbf{S}) &= \mathcal{N}(\mathbf{u}; \mathbf{m}, \mathbf{S})\frac{\partial}{\partial\mathbf{m}}\left[-\frac{1}{2}(\mathbf{u}-\mathbf{m})^\intercal\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})\right] \\ &= \mathcal{N}(\mathbf{u}; \mathbf{m}, \mathbf{S})\left(-\frac{1}{2}\right)(\mathbf{S}^{-1} + \mathbf{S}^{-\intercal})(\mathbf{u}-\mathbf{m})\frac{\partial}{\partial\mathbf{m}}[\mathbf{u}-\mathbf{m}] \\ &= \mathcal{N}(\mathbf{u}; \mathbf{m}, \mathbf{S})\frac{1}{2}(\mathbf{S}^{-1} + \mathbf{S}^{-\intercal})(\mathbf{u}-\mathbf{m}).\end{aligned}$$

Substituting it back into (6) gives

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{m}}\mathbb{E}[V_{\mathbf{r}}(s)] &= \frac{1}{2}\iiint V_{\mathbf{r}}(s)(\mathbf{S}^{-1}+\mathbf{S}^{-\intercal})(\mathbf{u}-\mathbf{m})p(\mathbf{r}|\boldsymbol{\lambda},\mathbf{X}_{\mathbf{u}},\mathbf{u})\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S})q(\boldsymbol{\lambda})\,d\mathbf{r}\,d\mathbf{u}\,d\boldsymbol{\lambda} \\
&= \frac{1}{2}\mathbb{E}[V_{\mathbf{r}}(s)(\mathbf{S}^{-1}+\mathbf{S}^{-\intercal})(\mathbf{u}-\mathbf{m})].
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{m}} =& -\frac{1}{2}(\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]+\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]^{\intercal})\mathbf{m}+\mathbf{t}^{\intercal}\mathbb{E}[\mathbf{K}_{\mathbf{r},\mathbf{u}}^{\intercal}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}] \\
&-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[V_{\mathbf{r}}(s_{i,t})(\mathbf{S}^{-1}+\mathbf{S}^{-\intercal})(\mathbf{u}-\mathbf{m})] \\
&-\gamma\sum_{s'\in\mathcal{S}}\mathcal{T}(s_{i,t},a_{i,t},s')\mathbb{E}[V_{\mathbf{r}}(s')(\mathbf{S}^{-1}+\mathbf{S}^{-\intercal})(\mathbf{u}-\mathbf{m})].
\end{aligned}$$

2.2 $\partial/\partial\mathbf{S}$

Similarly to the previous section,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{S}} =& \frac{1}{2}\frac{\partial}{\partial \mathbf{S}}\log|\mathbf{S}|-\frac{1}{2}\frac{\partial}{\partial \mathbf{S}}\operatorname{Tr}[\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]\mathbf{S}] \\
&-\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial}{\partial \mathbf{S}}\mathbb{E}[V_{\mathbf{r}}(s_{i,t})]-\gamma\sum_{s'\in\mathcal{S}}\mathcal{T}(s_{i,t},a_{i,t},s')\frac{\partial}{\partial \mathbf{S}}\mathbb{E}[V_{\mathbf{r}}(s')],
\end{aligned}$$

where

$$\frac{\partial}{\partial \mathbf{S}}\log|\mathbf{S}| = \mathbf{S}^{-\intercal},$$

and

$$\frac{\partial}{\partial \mathbf{S}}\operatorname{Tr}[\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]\mathbf{S}] = \mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]^{\intercal}$$

by *The Matrix Cookbook* [2]. Then

$$\frac{\partial}{\partial \mathbf{S}}\mathbb{E}[V_{\mathbf{r}}(s)] = \iiint V_{\mathbf{r}}(s)q(\mathbf{r})\frac{\partial}{\partial \mathbf{S}}\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S})q(\boldsymbol{\lambda})\,d\mathbf{r}\,d\mathbf{u}\,d\boldsymbol{\lambda},$$

where

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{S}}\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S}) &= \frac{\partial}{\partial \mathbf{S}}\left[\frac{1}{(2\pi)^{m/2}|\mathbf{S}|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})\right)\right] \\
&= \frac{\partial}{\partial \mathbf{S}}\left[\frac{1}{(2\pi)^{m/2}|\mathbf{S}|^{1/2}}\right]\exp\left(-\frac{1}{2}(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})\right) \\
&+ \frac{1}{(2\pi)^{m/2}|\mathbf{S}|^{1/2}}\frac{\partial}{\partial \mathbf{S}}\left[\exp\left(-\frac{1}{2}(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})\right)\right] \\
&= \frac{1}{(2\pi)^{m/2}}\frac{\partial}{\partial \mathbf{S}}\left[\frac{1}{|\mathbf{S}|^{1/2}}\right]\exp\left(-\frac{1}{2}(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})\right) \\
&- \frac{1}{2}\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S})\frac{\partial}{\partial \mathbf{S}}[(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})].
\end{aligned}$$

The two remaining derivatives can be taken with the help of *The Matrix Cookbook* [2]:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{S}}\left[\frac{1}{|\mathbf{S}|^{1/2}}\right] = -\frac{1}{2}|\mathbf{S}|^{-3/2}\frac{\partial|\mathbf{S}|}{\partial \mathbf{S}} = -\frac{1}{2}|\mathbf{S}|^{-3/2}|\mathbf{S}|\mathbf{S}^{-\intercal} &= -\frac{1}{2|\mathbf{S}|^{1/2}}\mathbf{S}^{-\intercal}, \\
\frac{\partial}{\partial \mathbf{S}}[(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-1}(\mathbf{u}-\mathbf{m})] &= -\mathbf{S}^{-\intercal}(\mathbf{u}-\mathbf{m})(\mathbf{u}-\mathbf{m})^{\intercal}\mathbf{S}^{-\intercal}.
\end{aligned}$$

Plugging them back in gives

$$\frac{\partial}{\partial \mathbf{S}}\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S}) = -\frac{1}{2}\mathbf{S}^{-\intercal}\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S}) + \frac{1}{2}\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S})\mathbf{S}^{-\intercal}(\mathbf{u}-\mathbf{m})(\mathbf{u}-\mathbf{m})^\intercal\mathbf{S}^{-\intercal},$$

and

$$\begin{aligned}\frac{\partial}{\partial \mathbf{S}}\mathbb{E}[V_\mathbf{r}(s)] &= \frac{1}{2}\iiint V_\mathbf{r}(s)(\mathbf{S}^{-\intercal}(\mathbf{u}-\mathbf{m})(\mathbf{u}-\mathbf{m})^\intercal\mathbf{S}^{-\intercal} - \mathbf{S}^{-\intercal})q(\mathbf{r})\mathcal{N}(\mathbf{u};\mathbf{m},\mathbf{S})q(\boldsymbol{\lambda})\,d\mathbf{r}\,d\mathbf{u}\,d\boldsymbol{\lambda} \\ &= \frac{1}{2}\mathbb{E}[V_\mathbf{r}(s)(\mathbf{S}^{-\intercal}(\mathbf{u}-\mathbf{m})(\mathbf{u}-\mathbf{m})^\intercal\mathbf{S}^{-\intercal} - \mathbf{S}^{-\intercal})].\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{S}} = \frac{1}{2}\mathbf{S}^{-\intercal} - \frac{1}{2}\mathbb{E}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}]^\intercal - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[V_\mathbf{r}(s_{i,t})(\mathbf{S}^{-\intercal}(\mathbf{u}-\mathbf{m})(\mathbf{u}-\mathbf{m})^\intercal\mathbf{S}^{-\intercal} - \mathbf{S}^{-\intercal})] \\ - \gamma\sum_{s'\in\mathcal{S}}\mathcal{T}(s_{i,t},a_{i,t},s')\mathbb{E}[V_\mathbf{r}(s')(\mathbf{S}^{-\intercal}(\mathbf{u}-\mathbf{m})(\mathbf{u}-\mathbf{m})^\intercal\mathbf{S}^{-\intercal} - \mathbf{S}^{-\intercal})].\end{aligned}$$

2.3 $\partial/\partial\alpha_j$

We begin in the usual way:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \alpha_j} =& -\frac{1}{2}\frac{\partial}{\partial \alpha_j}\mathbb{E}[\mathrm{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-2}]] - \frac{1}{2}\frac{\partial}{\partial \alpha_j}\mathbb{E}[\mathrm{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{S}]] - \frac{1}{2}\frac{\partial}{\partial \alpha_j}\mathbb{E}[\mathbf{m}^\intercal\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{m}] - \frac{1}{2}\frac{\partial}{\partial \alpha_j}\mathbb{E}[\log|\mathbf{K}_{\mathbf{u},\mathbf{u}}|] \\ &+ \frac{\partial}{\partial \alpha_j}\mathbb{E}[\mathbf{t}^\intercal\mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{m}] + \frac{\partial}{\partial \alpha_j}[\alpha_j + \log\Gamma(\alpha_j) + (1-\alpha_j)\psi(\alpha_j)] \\ &- \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial}{\partial \alpha_j}\mathbb{E}[V_\mathbf{r}(s_{i,t})] - \gamma\sum_{s'\in\mathcal{S}}\mathcal{T}(s_{i,t},a_{i,t},s')\frac{\partial}{\partial \alpha_j}\mathbb{E}[V_\mathbf{r}(s')].\end{aligned}$$

First,

$$\frac{\partial}{\partial \alpha_j}[\alpha_j + \log\Gamma(\alpha_j) + (1-\alpha_j)\psi(\alpha_j)] = 1 + \psi(\alpha_j) - \psi(\alpha_j) + (1-\alpha_j)\psi'(\alpha_j) = 1 + (1-\alpha_j)\psi'(\alpha_j)$$

by the definition of ψ. The remaining terms can all be treated in the same way, as they all contain expectations of scalar functions that are independent of α_j, and α_j only occurs in $\Gamma(\lambda_j;\alpha_j,\beta_j)$. Thus we can work with an abstract function as follows:

$$\begin{aligned}\frac{\partial}{\partial \alpha_j}\mathbb{E}[f(k_{\boldsymbol{\lambda}},\mathbf{r})] &= \frac{\partial}{\partial \alpha_j}\iiint f(k_{\boldsymbol{\lambda}},\mathbf{r})q(\boldsymbol{\lambda})q(\mathbf{r})q(\mathbf{u})\,d\boldsymbol{\lambda}\,d\mathbf{r}\,d\mathbf{u} \\ &= \iiint f(k_{\boldsymbol{\lambda}},\mathbf{r})q(\lambda_0)\cdots q(\lambda_{j-1})\frac{\partial}{\partial \alpha_j}\left[\frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\lambda_j^{\alpha_j-1}\right]e^{-\beta_j\lambda_j} \\ &\qquad\qquad q(\lambda_{j+1})\cdots q(\lambda_d)q(\mathbf{r})q(\mathbf{u})\,d\boldsymbol{\lambda}\,d\mathbf{r}\,d\mathbf{u}.\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial}{\partial \alpha_j}\left[\frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\lambda_j^{\alpha_j-1}\right] &= \frac{\frac{\partial}{\partial \alpha_j}[\beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}]\Gamma(\alpha_j) - \beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}\Gamma'(\alpha_j)}{(\Gamma(\alpha_j))^2} \\ &= \frac{\beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}\frac{\partial}{\partial \alpha_j}[\alpha_j\log\beta_j + (\alpha_j-1)\log\lambda_j]\Gamma(\alpha_j) - \beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}\Gamma'(\alpha_j)}{(\Gamma(\alpha_j))^2} \\ &= \frac{\beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}(\log\beta_j + \log\lambda_i)\Gamma(\alpha_j) - \beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}\Gamma'(\alpha_j)}{(\Gamma(\alpha_j))^2} \\ &= \frac{\beta_j^{\alpha_j}\lambda_j^{\alpha_j-1}}{\Gamma(\alpha_j)}\left(\log\beta_j + \log\lambda_j - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)}\right),\end{aligned}$$

which means that

$$\frac{\partial}{\partial \alpha_j}\mathbb{E}[f(k_{\boldsymbol{\lambda}}, \mathbf{r})] = \iiint f(k_{\boldsymbol{\lambda}}, \mathbf{r}) q(\lambda_0) \cdots q(\lambda_{j-1}) \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \lambda_j^{\alpha_j - 1} e^{-\beta_j \lambda_j} \left(\log \beta_j + \log \lambda_j - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} \right)$$
$$q(\lambda_{j+1}) \cdots q(\lambda_d) q(\mathbf{r}) q(\mathbf{u}) \, d\boldsymbol{\lambda} \, d\mathbf{r} \, d\mathbf{u}$$
$$= \iiint f(k_{\boldsymbol{\lambda}}, \mathbf{r}) \left(\log \beta_j + \log \lambda_j - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} \right) q(\boldsymbol{\lambda}) q(\mathbf{r}) q(\mathbf{u}) \, d\boldsymbol{\lambda} \, d\mathbf{r} \, d\mathbf{u}$$
$$= \mathbb{E}\left[f(k_{\boldsymbol{\lambda}}, \mathbf{r}) \left(\log \beta_j + \log \lambda_j - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} \right) \right]$$
$$= \left(\log \beta_j - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} \right) \mathbb{E}[f(k_{\boldsymbol{\lambda}}, \mathbf{r})] + \mathbb{E}[f(k_{\boldsymbol{\lambda}}, \mathbf{r}) \log \lambda_j].$$

With these results in mind, we can simplify the initial expression to

$$\frac{\partial \mathcal{L}}{\partial \alpha_j} = 1 + (1 - \alpha_j)\psi'(\alpha_j) + \left(\log \beta_j - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} \right) \Bigg(-\frac{1}{2}\mathbb{E}[\operatorname{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-2}]] - \frac{1}{2}\mathbb{E}[\operatorname{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{S}]]$$
$$- \frac{1}{2}\mathbb{E}[\mathbf{m}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{m}] - \frac{1}{2}\mathbb{E}[\log |\mathbf{K}_{\mathbf{u},\mathbf{u}}|] + \mathbb{E}[\mathbf{t}^\intercal \mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{m}]$$
$$- \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[V_{\mathbf{r}}(s_{i,t})] - \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s_{i,t}, a_{i,t}, s') \mathbb{E}[V_{\mathbf{r}}(s')] \Bigg)$$
$$- \frac{1}{2}\mathbb{E}[\operatorname{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-2}] \log \lambda_j] - \frac{1}{2}\mathbb{E}[\operatorname{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{S}] \log \lambda_j] - \frac{1}{2}\mathbb{E}[\mathbf{m}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{m} \log \lambda_j]$$
$$- \frac{1}{2}\mathbb{E}[\log |\mathbf{K}_{\mathbf{u},\mathbf{u}}| \log \lambda_j] + \mathbb{E}[\mathbf{t}^\intercal \mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{m} \log \lambda_j]$$
$$- \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[V_{\mathbf{r}}(s_{i,t}) \log \lambda_j] - \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s_{i,t}, a_{i,t}, s') \mathbb{E}[V_{\mathbf{r}}(s') \log \lambda_j].$$

2.4 $\partial/\partial\beta_j$

Finally,

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = -\frac{1}{2}\frac{\partial}{\partial \beta_j}\mathbb{E}[\operatorname{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-2}]] - \frac{1}{2}\frac{\partial}{\partial \beta_j}\mathbb{E}[\operatorname{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{S}]] - \frac{1}{2}\frac{\partial}{\partial \beta_j}\mathbb{E}[\mathbf{m}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{m}]$$
$$- \frac{1}{2}\frac{\partial}{\partial \beta_j}\mathbb{E}[\log |\mathbf{K}_{\mathbf{u},\mathbf{u}}|] + \frac{\partial}{\partial \beta_j}\mathbb{E}[\mathbf{t}^\intercal \mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1} \mathbf{m}] - \frac{\partial}{\partial \beta_j}[\log \beta_j]$$
$$- \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\partial}{\partial \beta_j}\mathbb{E}[V_{\mathbf{r}}(s_{i,t})] - \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s_{i,t}, a_{i,t}, s') \frac{\partial}{\partial \beta_j}\mathbb{E}[V_{\mathbf{r}}(s')].$$

Similarly to the previous section, we can handle all derivatives of expectations in the same way:

$$\frac{\partial}{\partial \beta_j}\mathbb{E}[f(k_{\boldsymbol{\lambda}}, \mathbf{r})] = \frac{\partial}{\partial \beta_j} \iiint f(k_{\boldsymbol{\lambda}}, \mathbf{r}) q(\boldsymbol{\lambda}) q(\mathbf{r}) q(\mathbf{u}) \, d\boldsymbol{\lambda} \, d\mathbf{r} \, d\mathbf{u}$$
$$= \iiint f(k_{\boldsymbol{\lambda}}, \mathbf{r}) q(\lambda_0) \cdots q(\lambda_{j-1}) \frac{\lambda_j^{\alpha_j - 1}}{\Gamma(\alpha_j)} \frac{\partial}{\partial \beta_j}[\beta_j^{\alpha_j} e^{-\beta_j \lambda_j}]$$
$$q(\lambda_{j+1}) \cdots q(\lambda_d) q(\mathbf{r}) q(\mathbf{u}) \, d\boldsymbol{\lambda} \, d\mathbf{r} \, d\mathbf{u}.$$

Since

$$\frac{\partial}{\partial \beta_j}[\beta_j^{\alpha_j} e^{-\beta_j \lambda_j}] = \alpha_j \beta_j^{\alpha_j - 1} e^{-\beta_j \lambda_j} - \beta_j^{\alpha_j} e^{-\beta_j \lambda_j} \lambda_j = \beta_j^{\alpha_j} e^{-\beta_j \lambda_j} \left(\frac{\alpha_j}{\beta_j} - \lambda_j \right),$$

we have that

$$
\begin{aligned}
\frac{\partial}{\partial \beta_j}\mathbb{E}[f(k_{\boldsymbol{\lambda}},\mathbf{r})] = & \iiint f(k_{\boldsymbol{\lambda}},\mathbf{r}) q(\lambda_0)\cdots q(\lambda_{j-1}) \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \lambda_j^{\alpha_j - 1} e^{-\beta_j \lambda_j} \left(\frac{\alpha_j}{\beta_j} - \lambda_j\right) \\
& \qquad q(\lambda_{j+1})\cdots q(\lambda_d) q(\mathbf{r}) q(\mathbf{u})\, d\boldsymbol{\lambda}\, d\mathbf{r}\, d\mathbf{u} \\
= & \iiint f(k_{\boldsymbol{\lambda}},\mathbf{r}) \left(\frac{\alpha_j}{\beta_j} - \lambda_j\right) q(\boldsymbol{\lambda}) q(\mathbf{r}) q(\mathbf{u})\, d\boldsymbol{\lambda}\, d\mathbf{r}\, d\mathbf{u} \\
= & \, \mathbb{E}\left[f(k_{\boldsymbol{\lambda}},\mathbf{r})\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] = \frac{\alpha_j}{\beta_j}\mathbb{E}[f(k_{\boldsymbol{\lambda}},\mathbf{r})] - \mathbb{E}[f(k_{\boldsymbol{\lambda}},\mathbf{r})\lambda_j].
\end{aligned}
$$

This gives us the final expression of $\frac{\partial \mathcal{L}}{\partial \beta_j}$:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta_j} = & -\frac{1}{\beta_j} - \frac{1}{2}\mathbb{E}\left[\mathrm{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-2}]\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] - \frac{1}{2}\mathbb{E}\left[\mathrm{Tr}[\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{S}]\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] \\
& - \frac{1}{2}\mathbb{E}\left[\mathbf{m}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{m}\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] - \frac{1}{2}\mathbb{E}\left[\log|\mathbf{K}_{\mathbf{u},\mathbf{u}}|\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] \\
& + \mathbb{E}\left[\mathbf{t}^\intercal \mathbf{K}_{\mathbf{r},\mathbf{u}}^\intercal \mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{m}\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] \\
& - \sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[V_{\mathbf{r}}(s_{i,t})\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right] - \gamma \sum_{s' \in \mathcal{S}} \mathcal{T}(s_{i,t}, a_{i,t}, s')\mathbb{E}\left[V_{\mathbf{r}}(s')\left(\frac{\alpha_j}{\beta_j} - \lambda_j\right)\right].
\end{aligned}
$$

References

[1] H. Herrlich. *Axiom of choice*. Springer, 2006.

[2] K. B. Petersen, M. S. Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

[3] H. Royden and P. Fitzpatrick. *Real Analysis*. Prentice Hall, 2010.