SUPPLEMENTARY MATERIAL — PROOFS

For any topological space X, let $\mathcal{K}(X)$ be the class of continuous real-valued functions from X having compact support: for any $f \in \mathcal{K}(X)$ there is some compact $K \subset X$ such that f is zero outside K. Any measure on X is always finite for any function on $\mathcal{K}(X)$ and to show that two measures are the same, it is sufficient that they agree for all functions in $\mathcal{K}(X)$.

Suppose from now on that Assumptions 1 and 2 hold: X is a topological space and G is a topological group acting continuously and *properly* on X, with both X and G Hausdorff and locally compact. Recall that the requirement that the action is proper means that the continuous function $\theta : X \times G \to X \times X$ defined by $(x,g) \mapsto (x, gx)$ is such that for any compact set $K \subset X \times X$, the pre-image $\theta^{-1}(K)$ of K is compact in $X \times G$.[6] For any $x \in X$, let $G_x := \{g \in G \mid gx = x\}$ be the *isotropy subgroup* of G at x and let $\pi_x : G \to G/G_x$ be the natural quotient map from G to the coset space G/G_x. Because G acts properly on X, each G_x is compact.

The image of $X \times G$ under θ is the set $E := \{(x, gx) \mid x \in X, g \in G\}$, which is closed in $X \times X$ because θ is a proper (hence closed) map and $X \times G$ is closed. If we restrict the codomain of θ to E, it becomes a surjective, continuous, and closed map: it is a quotient map. In other words, any set $U \subset E$ is open in the subspace topology inherited by E from $X \times X$ if and only if $\theta^{-1}(U)$ is open in $X \times G$. Further, θ has the following universal property: if Z is any topological space and $f : X \times G \to Z$ is a continuous function satisfying $f(x,g) = f(x',g')$ whenever $\theta(x,g) = \theta(x',g')$, then there is a unique continuous function $\bar{f} : E \to Z$ such that $f = \bar{f} \circ \theta$. We see that $\theta(x,g) = \theta(x',g')$ if and only if $x = x'$ and $g' \in gG_x$ (i.e., $gx = g'x$). The equivalence classes under θ are therefore sets of the form $\{x\} \times gG_x$.

Let λ be a χ-invariant measure on X under the action of G, where $\chi : G \to \mathbb{R}_+^\times$ is a continuous group homomorphism from G to the multiplicative group of the positive real numbers: for any measurable $F \subset X$ and $g \in G$, $\lambda(gF) = \chi(g)\lambda(F)$. Note that as a corollary we get that for any $f \in \mathcal{K}(X)$,

$$\int f(gx)\,\lambda(dx) = \chi(g^{-1}) \int f(x)\,\lambda(dx)\,. \tag{5}$$

Indeed, when $f = \chi_U$, $U \subset X$ measurable, $\int f(gx)\lambda(dx) = \int \mathbb{1}\{gx \in U\}\lambda(dx) = \int \mathbb{1}\{x \in g^{-1}U\}\lambda(dx) = \lambda(g^{-1}U) = \chi(g^{-1})\lambda(U) = \chi(g^{-1}) \int \mathbb{1}\{x \in U\}\lambda(dx) = \chi(g^{-1}) \int f(x)\lambda(dx)$, from which the result follows.

[6] More generally, $f : X \to Y$ is said to be proper if $f \otimes \mathrm{id}_Z : X \times Z \to Y \times Z$ is closed for every topological space Z, and a group is said to act properly if θ (as defined above) is proper. Our definition coincides with this one because the domain and codomain of θ are both locally compact.

Let μ be a left Haar measure on G. Recall that this means that $\mu(H) = \mu(gH)$ for any measurable $H \subset G$ and $g \in G$. We will also need the *right modular character* Δ_r^G of G. Recall that Δ_r^G is the unique function from G to the positive reals such that $\mu(Hg) = \Delta_r^G(g)\mu(H)$ for any measurable $H \subset G$. (The existence of Δ_r^G follows since $H \mapsto \mu(Hg)$ can be seen to be a left Haar measure on G and by the uniqueness of Haar measures up to a normalizing constant.) A well known fact, that we will need later, is that for any $f \in \mathcal{K}(G)$,

$$\int f(g^{-1})\,\mu(dg) = \int f(g)\Delta_r^G(g^{-1})\,\mu(dg)\,. \tag{6}$$

Finally, let β_x be a left Haar measure on G_x; by the compactness of G_x, β_x is also a right Haar measure and it is finite, and without loss of generality we can take it to be normalized.

For any $x \in X$ and $f \in \mathcal{K}(G)$, we will make use of the following construction: define $f'_x \in \mathcal{K}(G)$ by $g \mapsto \int_{G_x} f(gh)\,\beta_x(dh)$. For any $g' \in gG_x$, we have $f'_x(g') = \int_{G_x} f(gg^{-1}g'h)\,\beta_x(dh) = f'_x(g)$ since β_x is invariant under a translation by $g^{-1}g' \in G_x$. Thus f'_x is constant on each coset gG_x and there is some $f_x \in \mathcal{K}(G/G_x)$ such that $f'_x = f_x \circ \pi_x$. Because G_x is compact, there is a *quotient measure* $\nu_x := \mu/\beta_x$ on G/G_x which satisfies $\mu(f) = \nu_x(f_x)$ for any $f : \mathcal{K}(G)$. Furthermore, because β_x is normalized, $\nu_x = \pi_x(\mu)$.

Let M, N be measurable spaces, $\alpha : M \to N$ measurable, ρ a measure on M. The *push-forward measure* $\alpha(\rho)$ on N is defined by $\int f\, d\alpha(\rho) = \int f \circ \alpha\; d\rho$ for any $f \in \mathcal{K}(M)$ or by $\alpha(\rho)(F) = \rho(\alpha^{-1}(F))$ for any measurable $F \subset N$. From now on, $\alpha(\rho)$ for α an $M \to N$ map, ρ a measure on M always means the push-forward of ρ under α. In particular, the parentheses in a setting like this will never be used for grouping. To help parsing the formulae, we will also occasionally write $f \cdot \rho$ to denote the measure whose density w.r.t. ρ is f, where ρ is a measure on M and $f : M \to [0, \infty)$ is ρ-integrable.

Now consider a measure Γ on $X \times G$ defined by $\Gamma(dx, dg) := \gamma(x,g)\,\lambda(dx)\,\mu(dg)$, having density γ with respect to $\lambda \otimes \mu$. Our goal is to construct the Radon-Nikodym derivative of the push-forward measure $\theta(\Gamma)$ on E w.r.t. the push-forward measure $\theta(\lambda \otimes \mu)$. For this, take any $f \in \mathcal{K}(X \times G)$ so that

$$\begin{aligned}
\int f\, d\theta(\Gamma) &= \int f \circ \theta\, d\Gamma \\
&= \int_X \lambda(dx) \int_G \mu(dg)\,\gamma(x,g)\, f(\theta(x,g)) \\
&= \int_X \lambda(dx) \int_{G/G_x} \nu_x(dg) \int_{G_x} \beta_x(dh)\,\gamma(x,gh)\, f(\theta(x,gh))
\end{aligned}$$

$$= \int_X \lambda(dx) \int_{G/G_x} \nu_x(dg)\, f(\theta(x,g)) \int_{G_x} \beta_x(dh)\, \gamma(x,gh).$$

In the last equality, $f \circ \theta$ can be taken outside the innermost integral because $\theta(x,gh) = \theta(x,g)$ for any $h \in G_x$. Now define $\gamma'(x,g) := \int_{G_x} \beta_x(dh)\, \gamma(x,gh)$, so that $\gamma'(x,\cdot)$ is constant on each coset gG_x and there is some $\widetilde{\gamma} : E \to \mathbb{R}$ such that $\gamma' = \widetilde{\gamma} \circ \theta$:

$$\int f\, d\theta(\Gamma) = \int_X \lambda(dx) \int_{G/G_x} \nu_x(dg)\, f(\theta(x,g))\, \widetilde{\gamma}(\theta(x,g))\,.$$

The integrand of ν_x is well-defined because it depends on g only through its coset $\pi_x(g) = gG_x$. Using the fact that $\nu_x = \pi_x(\mu)$, we can replace ν_x by μ in the above integral to get

$$\begin{aligned} \int f\, d\theta(\Gamma) &= \int f(\theta(x,g))\, \widetilde{\gamma}(\theta(x,g))\, \lambda(dx)\, \mu(dg) \\ &= \int f \widetilde{\gamma}\, d\theta(\lambda \otimes \mu)\,. \end{aligned}$$

Thus we have shown that $\theta(\gamma \cdot (\lambda \otimes \mu)) = \widetilde{\gamma} \cdot \theta(\lambda \otimes \mu)$, where $\widetilde{\gamma}(\theta(x,g)) := \int_{G_x} \gamma(x,gh)\, \beta_x(dh)$.

We will be concerned with the operation of *transposition* on $X \times X$, defined by the map $(x,x')^T := T(x,x') = (x',x)$. We note that T is continuous and is its own inverse. Further, T maps the set E to itself: for any $(x,gx) \in E$ we have $T(x,gx) = (gx,x) = (gx, g^{-1}gx) \in E$. Mirroring this definition of T restricted to E, we will define $t : X \times G \to X \times G$ by $(x,g) \mapsto (gx, g^{-1})$, so that t is continuous and also its own inverse: $t(t(x,g)) = t(gx,g^{-1}) = (g^{-1}gx, g) = (x,g)$. Now note that if $\theta(x,g) = \theta(x,g')$ (i.e., $gx = g'x$) then $t(x,g) = (gx,g^{-1})$ and $t(x,g') = (g'x, g'^{-1})$, where $g'^{-1}g'x = x = g^{-1}gx$ and thus $\theta(t(x,g')) = \theta(t(x,g))$. Conversely, if $\theta(t(x,g)) = \theta(t(x',g'))$ then by the previous result $\theta(t(t(x,g))) = \theta(t(t(x',g')))$, and since t is its own inverse, we have shown that $\theta(t(x,g)) = \theta(t(x',g')) \iff \theta(x,g) = \theta(x',g')$. In other words, $\theta \circ t : X \times G \to E$ is constant on the equivalence classes of θ, so there is some continuous $\tau : E \to E$ such that $\theta \circ t = \tau \circ \theta$; we can verify that τ is simply T restricted to E, i.e., the following diagram is commutative:

$$\begin{array}{ccccc} X \times G & \xrightarrow{\ \theta\ } & E & \hookrightarrow & X \times X \\ \updownarrow{\scriptstyle t} & & \updownarrow{\scriptstyle T|_E} & & \updownarrow{\scriptstyle T} \\ X \times G & \xrightarrow{\ \theta\ } & E & \hookrightarrow & X \times X \end{array}$$

Let us again take $\Gamma = \gamma\,(\lambda \otimes \mu)$ and find the push-forward measure $t(\Gamma)$. Take $f \in \mathcal{K}(X \times G)$. Then,

$$\begin{aligned} \int f\, dt(\Gamma) &= \int f \circ t\, d\Gamma \\ &= \int f(gx, g^{-1})\, \gamma(x,g)\, \lambda(dx)\, \mu(dg) \end{aligned}$$

changing x to $g^{-1}x$ using Eq. (5)

$$= \int \chi(g^{-1})\, f(x,g^{-1})\, \gamma(g^{-1}x, g)\, \lambda(dx)\, \mu(dg)$$

changing g to g^{-1} using Eq. (6)

$$= \int \Delta_r^G(g^{-1})\, \chi(g)\, f(x,g)\, \gamma(gx, g^{-1})\, \lambda(dx)\, \mu(dg)\,.$$

Thus $t(\Gamma) = t(\gamma(\lambda \otimes \mu)) = \gamma_t\,(\lambda \otimes \mu)$ where $\gamma_t(x,g) := \varphi(g)\gamma(t(x,g))$ and $\varphi(g) = \chi(g)\Delta_r^G(g^{-1})$ for $g \in G$. Thus γ_t is a density for $t(\Gamma)$ with respect to $\lambda \otimes \mu$, so we can apply our previous result to this distribution to get a density for $\theta(t(\Gamma))$ with respect to $\theta(\lambda \otimes \mu)$: we get

$$\theta(t(\Gamma)) = \theta(\gamma_t\,(\lambda \otimes \mu)) = \widetilde{\gamma}_t \cdot \theta(\lambda \otimes \mu)\,,$$

where

$$\begin{aligned} \widetilde{\gamma}_t(\theta(x,g)) &:= \int_{G_x} \gamma_t(x,gh)\, \beta_x(dh) \\ &\overset{(a)}{=} \int_{G_x} \varphi(gh)\gamma(t(x,gh))\, \beta_x(dh) \\ &\overset{(b)}{=} \varphi(g) \int_{G_x} \gamma(ghx, h^{-1}g^{-1})\, \beta_x(dh) \\ &\overset{(c)}{=} \varphi(g) \int_{G_x} \gamma(gx, g^{-1}gh^{-1}g^{-1})\, \beta_x(dh) \\ &\overset{(d)}{=} \varphi(g) \int_{G_{gx}} \gamma(gx, g^{-1}h^{-1})\, \beta_{gx}(dh) \\ &\overset{(e)}{=} \varphi(g) \int_{G_{gx}} \gamma(gx, g^{-1}h)\, \beta_{gx}(dh) \\ &\overset{(f)}{=} \varphi(g)\widetilde{\gamma}(\theta(gx,g^{-1})) = \varphi(g)\widetilde{\gamma}(T(\theta(x,g)))\,. \end{aligned}$$

Here, the various equalities hold for the following reasons: (a) Definition of γ_t; (b) Since φ is a group homomorphism, $\varphi(gh) = \varphi(g)\varphi(h)$ and since G_x is compact, $\varphi(h) = 1$ for any $h \in G_x$; (c) By the definition of G_x, $hx = x$; (d) β_{gx} is the push-forward of β_x under the map $c_g : h \mapsto ghg^{-1}$. Indeed, if $\hat{\beta} := c_g(\beta_x)$ then $\hat{\beta}(U) = \beta_x(g^{-1}Ug)$ for $U \subset G_{gx}$ measurable. Now, for any $h \in G_{gx}$, $hU = U$, hence $\hat{\beta}(hU) = \beta_x(g^{-1}hUg) = \beta_x(g^{-1}Ug) = \hat{\beta}(U)$ and thus $\hat{\beta} = c_g(\beta_x)$ is a Haar-measure on G_{gx}. Thanks to the uniqueness of normalized Haar measures, we then have $c_g(\beta_x) = \beta_{gx}$; (e) Since G_{gx} is compact, β_{gx} remains unchanged under the change of variables $h \mapsto h^{-1}$; (f) Definition of $\widetilde{\gamma}$.

Theorem 3. *Let X, G, λ, μ, $(G_x)_{x \in X}$, $(\beta_x)_{x \in X}$ be as stated in this section. Then, for any Γ measure on $X \times G$ that is absolute continuous w.r.t. $\lambda \otimes \mu$, with density γ, it holds that*

$$\frac{d\theta(\Gamma)}{dT(\theta(\Gamma))}(x,gx) = \frac{\Delta_r^G(g)\, \widetilde{\gamma}(x,gx)}{\chi(g)\, \widetilde{\gamma}(gx,x)} \quad \text{where } x \in X, g \in G\,,$$

where $\theta(x,g) = (x,gx)$ and $T(x,x') = (x',x)$ for any $x,x' \in X$, $g \in G$ and

$$\widetilde{\gamma}(x,gx) = \int_{G_x} \gamma(x,gh)\, \beta_x(dh) \qquad \text{where } x \in X, g \in G\,.$$

Proof. $\varphi(\widetilde{\gamma} \circ T)$ is a density for $\theta(t(\Gamma))$ (and hence for $T(\theta(\Gamma))$ with respect to $\theta(\lambda \otimes \mu)$). Since the density for $\theta(\Gamma)$ with respect to the same measure is $\widetilde{\gamma}$, we see that the Radon-Nikodym derivative $d\theta(\Gamma)/dT(\theta(\Gamma))$ is $\widetilde{\gamma}(x, gx)/\varphi(g)\widetilde{\gamma}(gx, x)$ at $(x, gx) \in E$. □

We will now restate some results of Tierney (1998) for use in the following proofs.

Proposition 2 (Tierney, 1998, Proposition 1)**.** *Let $\mu(dx, dy)$ be a sigma-finite measure on the product space $(E \times E, \mathscr{E} \otimes \mathscr{E})$ and let $\mu^T(dx, dy) = \mu(dy, dx)$. Then there exists a symmetric set $R \in \mathscr{E} \otimes \mathscr{E}$ such that μ and μ^T are mutually absolutely continuous on R and mutually singular on the complement of R, R^C. The set R is unique up to sets that are null for both μ and μ^T. Let μ_R and μ_T^T be the restrictions of μ and μ^T to R. Then there exists a version of the density*

$$r(x, y) = \frac{\mu_R(dx, dy)}{\mu_R^T(dx, dy)}$$

such that $0 < r(x, y) < \infty$ and $r(x, y) = 1/r(y, x)$ for all $x, y \in E$.

Proposition 3 (Tierney, 1998, Theorem 2)**.** *A Metropolis-Hastings transition kernel satisfies the detailed balance condition Eq.* (1) *if and only if the following two conditions hold.*

(i) The function α is μ-almost everywhere zero on R^C.

(ii) The function α satisfies $\alpha(x, y)r(x, y) = \alpha(y, x)$ μ-almost everywhere on R.

The Metropolis-Hastings acceptance probability

$$\alpha(x, y) = \begin{cases} \min\{1, r(y, x)\}, & \text{if } (x, y) \in R, \\ 0, & \text{if } (x, y) \notin R. \end{cases}$$

satisfies these conditions by construction.

Proofs of Theorems 1 and 2

Proof of Theorem 1. Procedure 1 describes an MH kernel based on the proposal $Q'(dw' \mid w)$ that, given a state w, samples $g \sim Q_G(\cdot \mid w)$ and proposes gw. In other words, $Q'(\cdot \mid w)$ is the push-forward of $Q_G(\cdot \mid w)$ under the map $g \mapsto gw$, making $P(dw)\, Q'(dw' \mid w)$ the push-forward of $P(dw)\, Q_G(dg \mid w)$ under the map $\theta(w, g) = (w, gw)$. We can now apply Theorem 3 by taking $\Gamma(dw, dg) := P(dw)\, Q_G(dg \mid w)$ with density $\gamma(w, g) = p(w)\, q(g \mid w)$, so that $P(dw)\, Q'(dw' \mid w) = \theta(\Gamma)$ and

$$\begin{aligned} r(w, gw) &:= \frac{d\theta(P(dw)\, Q_G(dg \mid w))}{dT(\theta(P(dw)\, Q_G(dg \mid w)))}(w, gw) \\ &= \frac{\Delta_r^G(g)\, \widetilde{\gamma}(w, gw)}{\chi(g)\, \widetilde{\gamma}(gw, w)} \qquad \text{where } w \in W, g \in G \end{aligned}$$

where

$$\begin{aligned} \widetilde{\gamma}(w, gw) &= \int_{G_x} p(w)\, q(gh \mid w)\, \beta_x(dh) \\ &= p(w) \int_{G_x} q(gh \mid w)\, \beta_x(dh) \\ &= p(w)\, q'(g \mid w). \end{aligned}$$

Define

$$R := \big\{(w, gw) \in E \,\big|\, p(w)\, q'(g \mid w) > 0 \text{ and } p(gw)\, q'(g^{-1} \mid gw) > 0\big\}.$$

Now the image of θ is E, so both $\theta(\Gamma)$ and $T(\theta(\Gamma))$ are zero outside E. Thus they are mutually singular outside $R \subset E$ and mutually absolutely continuous on R. We can define $r(w, w') = 1$ outside R, and by inspection we can verify that $r(w', w) = 1/r(w, w')$. Thus we have satisfied all the conditions for Proposition 2 and by Proposition 3 the MH kernel with acceptance probability $\alpha(w, w') := \min\{1, r(w', w)\}$ on R satisfies detailed balance. Since we assume that the initial state is within the support of P, and the acceptance probability is always zero for proposals outside the support, α will never be evaluated outside the set R. □

Proof of Theorem 2. Procedure 2 describes an MH kernel based on a proposal Q' which is a mixture of the types of proposals seen in Procedure 1: $Q'(dw' \mid w) = \sum_{i=1}^n a(i \mid w) Q_i'(dw' \mid w)$ and $P(dw)\, Q'(dw' \mid w) = \sum_{i=1}^n a(i \mid w)\, P(dw)\, Q_i'(dw' \mid w)$. Now define $\Gamma_i(dw, dg) = a(i \mid w)\, P(dw)\, Q_i(dg \mid w)$. By a similar argument to the previous proof it follows that $P(dw)\, Q'(dw' \mid w) = \sum_{i=1}^n \theta(\Gamma_i)$. As before, we can define a function r_i that is a Radon-Nikodym derivative for $d\theta(\Gamma_i)/dT(\theta(\Gamma_i))$ restricted to a set R_i where both those measures are mutually absolutely continuous, and mutually singular outside it. Since $\theta(\Gamma_i)$ is zero outside the set $E_i := \theta(W, G_i)$, we see that $R_i \subset E_i$. The problem arises because the E_i may not be disjoint. However, we will show that we can take the R_i to be disjoint without loss of generality.

For each $1 \le i \le n$, define V_i to contain all the $1 \le j \le n$ such that Assumption 3 is satisfied for i and j with $k = i$. Now for any $j \in V_i$ the pre-image of $E_i \cap E_j$ under θ is $\big\{(w, g) \,\big|\, w \in W, g \in G_{i,j} G_i, w\big\}$. Applying the assumption, this set has zero measure under Γ_i so $E_i \cap E_j$ has zero measure under $\theta(\Gamma_i)$. Then $\bigcup_{j \in V_i} E_i \cap E_j$ has zero measure under $\theta(\Gamma_i)$ and is symmetric, so it has zero measure under $T(\theta(\Gamma_i))$ as well. Thus, without loss of generality, we can take R_i to be a subset of $E_i \setminus \bigcup_{j \in V_i} E_j$ since it is only unique up to $\theta(\Gamma_i)$-null sets. By the assumption, for any $i \neq j$ either $i \in V_j$ or $j \in V_i$, so the R_i are disjoint. We have found a collection of disjoint sets R_i such that each $\theta(\Gamma_i)$ is mutually absolutely continuous on R_i and mutually singular outside R_i, with $d\theta(\Gamma_i)/d(T(\theta(\Gamma_i))) = r_i$ restricted to

R_i. We can now define r so that it takes on the value r_i on E_i, with $R \coloneqq \bigcup_{i=1}^{n} R_i$. This r is the Radon-Nikodym derivative for Tierney's Proposition 1.

It only remains to note that by Assumption 3 for any w in the support of P and $w' = gw$ sampled according to $Q_i(\cdot \mid w)$, $(w, gw) \in R_i$ with probability 1. Thus if the algorithm samples from some Q_i then r is evaluated on E_i with probability 1. $\square$