Appendices

A Geometric Dynamical Systems

Here we summarize details and properties of GDSs introduced in Section 4.1.

A.1 From Geometric Mechanics to GDSs

Our study of GDSs is motivated by geometric mechanics. Many formulations of mechanics exist, including Lagrangian mechanics [35] and the aforementioned Gauss's Principle of Least Constraint [36]—They are all equivalent, implicitly sharing same mathematical structure. In that sense, geometric mechanics, which models physical systems as geodesic flow on Riemannian manifolds, is the most explicit of these, revealing directly the underlying manifold structure and connecting to the broad mathematical tool set from Riemannian geometry. These connections enable us here to generalize beyond the previous simple mechanical systems studied in [15] to non-classical systems that more naturally describe robotic behaviors with non-Euclidean geometric properties.

A.2 Degenerate GDSs

Let us recall the definition of GDSs.

Definition 1. *Let* $\mathbf{B} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_+^{m\times m}$ *and let* $\mathbf{G} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_+^{m\times m}$ *and* $\Phi : \mathbb{R}^m \to \mathbb{R}$ *be differentiable. We say the tuple* $(\mathcal{M}, \mathbf{G}, \mathbf{B}, \Phi)$ *is a* GDS *if*

$$\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}})\ddot{\mathbf{x}} + \boldsymbol{\xi}_{\mathbf{G}}(\mathbf{x}, \dot{\mathbf{x}}) = -\nabla_{\mathbf{x}}\Phi(\mathbf{x}) - \mathbf{B}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} \tag{5}$$

where $\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{G}(\mathbf{x}, \dot{\mathbf{x}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{x}, \dot{\mathbf{x}})$.

For degenerate cases, $\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}})$ can be singular and (5) define rather a family of differential equations. Degenerate cases are not uncommon; for example, the leaf-node dynamics could have $\mathbf{G}$ being only positive semidefinite. Having degenerate GDSs does not change the properties that we have proved, but one must be careful about whether differential equation satisfying (5) exist. For example, the existence is handled by the assumption on $\mathbf{M}$ in Theorem 1 and the assumption on $\mathbf{M}_r$ in Corollary 1. For RMPflow, we only need that $\mathbf{M}_r$ at the root node is non-singular. In other words, the natural-form RMP created by `pullback` at the root node can be resolved in the canonical-form RMP for policy execution. A sufficient and yet practical condition is provided in Theorem 2.

A.3 Geodesic and Stability

For GDSs, they possess a natural conservation property of kinematic energy, i.e. it travels along a geodesic defined by $\mathbf{G}(\mathbf{x}, \dot{\mathbf{x}})$ when there is no external perturbations due to Φ and $\mathbf{B}$. Note $\mathbf{G}(\mathbf{x}, \dot{\mathbf{x}})$ by definition may only be positive-semidefinite even when the system is non-degenerate; here we allow the geodesic

to be defined for a degenerate metric, meaning a curve whose instant length measured by the (degenerate) metric is constant.

This geometric feature is an important tool to establish the stability of non-degenerate GDSs; We highlight this nice geometric property below, which is a corollary of Proposition 1.

Corollary 3. *All non-degenerate GDSs in the form* $(\mathcal{M}, \mathbf{G}, 0, 0)$ *travel on geodesics. That is,* $K(\mathbf{x}, \dot{\mathbf{x}}) = 0$*, where* $K(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}\dot{\mathbf{x}}^\top \mathbf{G}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}}$.

Note that this property also hold for degenerate GDSs provided that differential equations satisfying (5) exist.

A.4 Curvature Term and Coriolis Force

The curvature term $\boldsymbol{\xi}_{\mathbf{G}}$ in GDSs is highly related to the Coriolis force in the mechanics literature. This is not surprising, as from the analysis in Section 4.4 we know that $\boldsymbol{\xi}_{\mathbf{G}}$ comes from the Christoffel symbols of the asymmetric connection. Recall it is defined as

$$\boldsymbol{\xi}_{\mathbf{G}}(\mathbf{x}, \dot{\mathbf{x}}) \coloneqq \overset{\mathbf{x}}{\mathbf{G}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} - \tfrac{1}{2}\nabla_{\mathbf{x}}(\dot{\mathbf{x}}^\top \mathbf{G}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}})$$

We show their relationship explicitly below.

Lemma 1. *Let* $\Gamma_{ijk} = \frac{1}{2}(\partial_{x_k} G_{ij} + \partial_{x_j} G_{ik} - \partial_{x_j} G_{jk})$ *be the Christoffel symbol of the first kind with respect to* $\mathbf{G}(\mathbf{x}, \dot{\mathbf{x}})$*, where the subscript* $_{ij}$ *denotes the* (i, j) *element. Let* $C_{ij} = \sum_{k=1}^{d} \dot{x}_k \Gamma_{ijk}$ *and define* $\mathbf{C}(\mathbf{x}, \dot{\mathbf{x}}) = (C_{ij})_{i,j=1}^{m}$*. Then* $\boldsymbol{\xi}_{\mathbf{G}}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{C}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}}$.

Proof of Lemma 1. Suppose $\boldsymbol{\xi}_{\mathbf{G}} = (\xi_i)_{i=1}^{m}$. We can compare the two definitions and verify they are indeed equivalent:

$$\begin{aligned}
\xi_i &= \sum_{j,k=1}^{d} \dot{x}_j \dot{x}_k \partial_{x_j} G_{ik} - \frac{1}{2} \sum_{j,k=1}^{d} \dot{x}_j \dot{x}_k \partial_{x_i} G_{jk} \\
&= \frac{1}{2} \sum_{j,k=1}^{d} \dot{x}_j \dot{x}_k \partial_{x_k} G_{ij} + \frac{1}{2} \sum_{j,k=1}^{d} \dot{x}_j \dot{x}_k \partial_{x_j} G_{ik} - \frac{1}{2} \sum_{j,k=1}^{d} \dot{x}_j \dot{x}_k \partial_{x_i} G_{jk} \\
&= (\mathbf{C}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}})_i
\end{aligned}$$

■

B Proofs of RMPflow Analysis

B.1 Proof of Theorem 1

Theorem 1. *Let the* i*th child node follow* $(\mathcal{N}_i, \mathbf{G}_i, \mathbf{B}_i, \Phi_i)_{\mathcal{S}_i}$ *and have coordinate* $\mathbf{y}_i$*. Let* $\mathbf{f}_i = -\boldsymbol{\eta}_{\mathbf{G}_i;\mathcal{S}_i} - \nabla_{\mathbf{y}_i}\Phi_i - \mathbf{B}_i\dot{\mathbf{y}}_i$ *and* $\mathbf{M}_i = \mathbf{G}_i + \boldsymbol{\Xi}_{\mathbf{G}_i}$*. If* $[\mathbf{f}, \mathbf{M}]^{\mathcal{M}}$ *of the parent node is given by* `pullback` *with* $\{[\mathbf{f}_i, \mathbf{M}_i]^{\mathcal{N}_i}\}_{i=1}^{K}$ *and* $\mathbf{M}$ *is non-singular, the parent node follows* $(\mathcal{M}, \mathbf{G}, \mathbf{B}, \Phi)_{\mathcal{S}}$*, where* $\mathbf{G} = \sum_{i=1}^{K} \mathbf{J}_i^\top \mathbf{G}_i \mathbf{J}_i$*,* $\mathbf{B} = \sum_{i=1}^{K} \mathbf{J}_i^\top \mathbf{B}_i \mathbf{J}_i$*,* $\Phi = \sum_{i=1}^{K} \Phi_i \circ \mathbf{y}_i$*,* $\mathcal{S}$ *preserves* $\mathcal{S}_i$*, and* $\mathbf{J}_i = \partial_{\mathbf{x}} \mathbf{y}_i$*. Particularly, if* $\mathbf{G}_i$ *is velocity-free and the child nodes are GDSs, the parent node follows* $(\mathcal{M}, \mathbf{G}, \mathbf{B}, \Phi)$.

Proof of Theorem 1. We will use the non-degeneracy assumption that $\mathbf{G} + \mathbf{\Xi}_{\mathbf{G}}$ (i.e. $\mathbf{M}$ as we will show) is non-singular, so that the differential equation specified by an RMP in normal form or a (structured) GDS is unique. This assumption is made to simplify writing. At the end of the proof, we will show that this assumption only needs to be true at the root node of RMPflow.

The general case We first show the differential equation given by `pullback` is equivalent to the differential equation of pullback structured GDS $(\mathcal{M}, \mathbf{G}, \mathbf{B}, \Phi)_{\mathcal{S}}$. Under the non-degeneracy assumption, suppose $\mathcal{S}_i$ factorizes $\mathbf{G}_i$ as $\mathbf{G}_i = \mathbf{L}_i^\top \mathbf{H}_i \mathbf{L}_i$, where $\mathbf{L}_i$ is some Jacobian matrix. On one hand, for `pullback`, because in the child node $\ddot{\mathbf{y}}_i$ satisfies $(\mathbf{G}_i + \mathbf{\Xi}_{\mathbf{G}_i})\ddot{\mathbf{y}}_i = -\boldsymbol{\eta}_{\mathbf{G}_i;\mathcal{S}_i} - \nabla_{\mathbf{y}_i}\Phi_i - \mathbf{B}_i\dot{\mathbf{y}}_i$ (where by definition $\boldsymbol{\eta}_{\mathbf{G}_i;\mathcal{S}_i} = \mathbf{L}_i^\top(\boldsymbol{\xi}_{\mathbf{H}_i} + (\mathbf{H}_i + \mathbf{\Xi}_{\mathbf{H}_i})\dot{\mathbf{L}}_i\dot{\mathbf{y}}_i)$), the `pullback` operator combines the child nodes into the differential equation at the parent node,

$$\mathbf{M}\ddot{\mathbf{x}} = \sum_{i=1}^{K} \mathbf{J}_i^\top \mathbf{M}_i(\ddot{\mathbf{y}}_i - \dot{\mathbf{J}}_i\dot{\mathbf{x}}) \tag{6}$$

where we recall $\mathbf{M} = \sum_{i=1}^{K} \mathbf{J}_i^\top \mathbf{M}_i \mathbf{J}_i$ is given by `pullback`. On the other hand, for $(\mathcal{M}, \mathbf{G}, \mathbf{B}, \Phi)_{\mathcal{S}}$ with $\mathcal{S}$ preserving $\mathcal{S}_i$, its dynamics satisfy

$$(\mathbf{G} + \mathbf{\Xi}_{\mathbf{G}})\,\ddot{\mathbf{x}} + \boldsymbol{\eta}_{\mathbf{G};\mathcal{S}} = -\nabla_{\mathbf{x}}\Phi - \mathbf{B}\dot{\mathbf{x}} \tag{7}$$

where $\mathbf{G}$ is factorized by $\mathcal{S}$ into

$$\mathbf{G} = \begin{bmatrix} \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_K \end{bmatrix}^\top \begin{bmatrix} \mathbf{G}_1 & & \\ & \ddots & \\ & & \mathbf{G}_K \end{bmatrix} \begin{bmatrix} \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_K \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1\mathbf{J}_1 \\ \vdots \\ \mathbf{L}_K\mathbf{J}_K \end{bmatrix}^\top \begin{bmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_K \end{bmatrix} \begin{bmatrix} \mathbf{L}_1\mathbf{J}_1 \\ \vdots \\ \mathbf{L}_K\mathbf{J}_K \end{bmatrix} =: \bar{\mathbf{J}}^\top \bar{\mathbf{H}} \bar{\mathbf{J}}$$

and the curvature term $\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}$ by $\mathcal{S}$ is given as $\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}} \coloneqq \bar{\mathbf{J}}^\top(\boldsymbol{\xi}_{\bar{\mathbf{H}}} + (\bar{\mathbf{H}} + \mathbf{\Xi}_{\bar{\mathbf{H}}})\dot{\bar{\mathbf{J}}}\dot{\mathbf{x}})$.

To prove the general statement, we will show (6) and (7) are equivalent. First, we introduce a lemma to write $\mathbf{\Xi}_{\mathbf{G}}$ in terms of $\mathbf{\Xi}_{\mathbf{G}_i}$ (proved later in this section).

Lemma 2. *Let $\mathcal{M}$ and $\mathcal{N}$ be two manifolds and let $\mathbf{x}$ and $\mathbf{y}(\mathbf{x})$ be the coordinates. Define $\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{J}(\mathbf{x})^\top \mathbf{N}(\mathbf{y}, \dot{\mathbf{y}})\mathbf{J}(\mathbf{x})$, where $\mathbf{J}(\mathbf{x}) = \partial_{\mathbf{x}}\mathbf{y}(\mathbf{x})$. Then*

$$\mathbf{\Xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{J}^\top(\mathbf{x})\mathbf{\Xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})\mathbf{J}(\mathbf{x})$$

Therefore, we see that on the LHSs

$$(\mathbf{G} + \mathbf{\Xi}_{\mathbf{G}})\ddot{\mathbf{x}} = \mathbf{M}\ddot{\mathbf{x}}$$

and on the RHSs

$$
\begin{aligned}
&\left(\sum_{i=1}^{K} \mathbf{J}_i^\top \mathbf{M}_i(\ddot{\mathbf{y}}_i - \dot{\mathbf{J}}_i\dot{\mathbf{x}})\right) \\
&= \left(\sum_{i=1}^{K} \mathbf{J}_i^\top (-\boldsymbol{\eta}_{\mathbf{G}_i;\mathcal{S}_i} - \nabla_{\mathbf{y}_i}\Phi_i - \mathbf{B}_i\dot{\mathbf{y}}_i - (\mathbf{G}_i + \boldsymbol{\Xi}_{\mathbf{G}_i})\dot{\mathbf{J}}_i\dot{\mathbf{x}})\right) \\
&= \left(\sum_{i=1}^{K} \mathbf{J}_i^\top (-\mathbf{L}_i^\top(\boldsymbol{\xi}_{\mathbf{H}_i} + (\mathbf{H}_i + \boldsymbol{\Xi}_{\mathbf{H}_i})\dot{\mathbf{L}}_i\dot{\mathbf{y}}_i) - (\mathbf{G}_i + \boldsymbol{\Xi}_{\mathbf{G}_i})\dot{\mathbf{J}}_i\dot{\mathbf{x}})\right) + \left(\sum_{i=1}^{K} \mathbf{J}_i^\top(-\nabla_{\mathbf{y}_i}\Phi_i - \mathbf{B}_i\dot{\mathbf{y}}_i)\right) \\
&= \left(\sum_{i=1}^{K} -\bar{\mathbf{J}}_i^\top \boldsymbol{\xi}_{\mathbf{H}_i} - \bar{\mathbf{J}}_i^\top(\mathbf{H}_i + \boldsymbol{\Xi}_{\mathbf{H}_i})(\dot{\mathbf{L}}_i\mathbf{J}_i + \mathbf{L}_i\dot{\mathbf{J}}_i)\dot{\mathbf{x}}\right) - \nabla_{\mathbf{x}}\Phi - \mathbf{B}\dot{\mathbf{x}} \\
&= -\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}} - \nabla_{\mathbf{x}}\Phi - \mathbf{B}\dot{\mathbf{x}}
\end{aligned}
$$

where the first equality is due to Lemma 2, the second equality is due to (6), and the third equality is due to the definition of structured GDSs. The above derivations show the equivalence between the RHSs and LHSs of (6) and (7), respectively. Therefore, when the non-degenerate assumption holds, (6) and (7) are equivalent.

The special case With the closure of structured GDSs proved, we next show the closure of GDSs under `pullback`, when the metric is only configuration-dependent. That is, we want to show that, when the metric is only configuration-dependent, the choice of structure does not matter. This amounts to show that $\boldsymbol{\xi}_{\mathbf{G}} = \boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}$ because by definition $\boldsymbol{\Xi}_i = 0$ and $\boldsymbol{\Xi} = 0$. Below we show how $\boldsymbol{\xi}_{\mathbf{G}}$ is written in terms of $\boldsymbol{\xi}_{\mathbf{G}_i}$ and $\boldsymbol{\Xi}_{\mathbf{G}_i}$ for general metric matrices and specialize it to the configuration-dependent special case (proved later in this section).

Lemma 3. *Let $\mathcal{M}$ and $\mathcal{N}$ be two manifolds and $\mathbf{x}$ and $\mathbf{y}(\mathbf{x})$ be the coordinates. Suppose $\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}})$ is structured as $\mathbf{J}(\mathbf{x})^\top\mathbf{N}(\mathbf{y}, \dot{\mathbf{y}})\mathbf{J}(\mathbf{x})$, where $\mathbf{J}(\mathbf{x}) = \partial_{\mathbf{x}}\mathbf{y}(\mathbf{x})$. Then*

$$
\begin{aligned}
\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{J}(\mathbf{x})^\top &\left(\boldsymbol{\xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}}) + (\mathbf{N}(\mathbf{y}, \dot{\mathbf{y}}) + 2\boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}}))\dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}}\right) \\
&- \dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})^\top\boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})^\top\mathbf{J}(\mathbf{x})\dot{\mathbf{x}}
\end{aligned}
$$

When $\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{M}(\mathbf{x})$, $\boldsymbol{\xi}_{\mathbf{M}} = \boldsymbol{\eta}_{\mathbf{M};\mathcal{S}}$ regardless of the structure of $\mathcal{S}$.

By Lemma 3, we see that structured GDSs are GDSs regardless of the chosen structure when the metric is only configuration dependent. Thus, the statement of the special case follows by combining Lemma 3 and the previous proof for structured GDSs.

Remarks: Proof of Corollary 1 We note that the non-degenerate assumption does not need to hold for every nodes in RMPflow but only for the root node. This can be seen from the proof above, where we propagate the LHSs and RHSs *separately*. Therefore, as long as the inertial matrix at the root node is invertible, the differential equation on the configuration space is well defined. ■

Proof of Lemma 2. Let $\mathbf{m}_i$, $\mathbf{n}_i$, $\mathbf{j}_i$ be the ith column of $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{J}$, respectively. Suppose $\mathcal{M}$ and $\mathcal{N}$ are of m and n dimensions, respectively. By definition of $\boldsymbol{\Xi}_{\mathbf{M}}$,

$$\begin{aligned}
2\boldsymbol{\Xi}_{\mathbf{M}}(\mathbf{x},\dot{\mathbf{x}}) = \sum_{i=1}^{m} \dot{x}_i \partial_{\dot{\mathbf{x}}} \mathbf{m}_i(\mathbf{x},\dot{\mathbf{x}}) &= \mathbf{J}(\mathbf{x})^\top \sum_{i=1}^{m} \dot{x}_i \partial_{\dot{\mathbf{x}}} (\mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\mathbf{j}_i(\mathbf{x})) \\
&= \mathbf{J}(\mathbf{x})^\top \left(\sum_{i=1}^{m} \dot{x}_i \partial_{\dot{\mathbf{y}}} (\mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\mathbf{j}_i(\mathbf{x})) \right) \mathbf{J}(\mathbf{x}) \\
&= \mathbf{J}(\mathbf{x})^\top \left(\sum_{j=1}^{n} \partial_{\dot{\mathbf{y}}} \mathbf{n}_j(\mathbf{y},\dot{\mathbf{y}}) \sum_{i=1}^{m} \dot{x}_i J_{ji}(\mathbf{x}) \right) \mathbf{J}(\mathbf{x}) \\
&= \mathbf{J}(\mathbf{x})^\top \left(\sum_{j=1}^{n} y_j \partial_{\dot{\mathbf{y}}} \mathbf{n}_j(\mathbf{y},\dot{\mathbf{y}}) \right) \mathbf{J}(\mathbf{x}) \\
&= 2\mathbf{J}(\mathbf{x})^\top \boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}}) \mathbf{J}(\mathbf{x}) \qquad \blacksquare
\end{aligned}$$

Proof of Lemma 3. Before the proof, we first note a useful identity $\partial_{\mathbf{x}} \dot{\mathbf{y}} = \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})$. This can be derived simply by the definition of the Jacobian matrix $(\partial_{\mathbf{x}} \mathbf{J}(\mathbf{x})\dot{\mathbf{x}})_{ij} = \sum_{k=1}^{m} \dot{x}_k \partial_{x_j} J_{ik} = \sum_{k=1}^{m} \dot{x}_k \partial_{x_j} \partial_{x_k} y_i = \sum_{k=1}^{m} \dot{x}_k \partial_{x_k} J_{ij} = (\dot{\mathbf{J}})_{ij}$.

To prove the lemma, we derive $\boldsymbol{\xi}_{\mathbf{M}}$ by its definition

$$\begin{aligned}
\boldsymbol{\xi}_{\mathbf{M}} &= \overset{\times}{\mathbf{M}}(\mathbf{x},\dot{\mathbf{x}})\dot{\mathbf{x}} - \frac{1}{2}\nabla_{\mathbf{x}}(\dot{\mathbf{x}}^\top \mathbf{M}(\mathbf{x},\dot{\mathbf{x}})\dot{\mathbf{x}}) \\
&= \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\mathbf{J}(\mathbf{x})\dot{\mathbf{x}} + \mathbf{J}(\mathbf{x})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})\dot{\mathbf{x}} + \mathbf{J}(\mathbf{x})^\top \overset{\times}{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})\mathbf{J}(\mathbf{x})\dot{\mathbf{x}} - \frac{1}{2}\nabla_{\mathbf{x}}(\dot{\mathbf{x}}^\top \mathbf{J}(\mathbf{x})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\mathbf{J}(\mathbf{x})\dot{\mathbf{x}}) \\
&= \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} + \mathbf{J}(\mathbf{x})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})\dot{\mathbf{x}} + \mathbf{J}(\mathbf{x})^\top \overset{\times}{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} - \frac{1}{2}\nabla_{\mathbf{x}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) \\
&= \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} + \mathbf{J}(\mathbf{x})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})\dot{\mathbf{x}} + \mathbf{J}(\mathbf{x})^\top \overset{\times}{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} \\
&\quad - \frac{1}{2}\mathbf{J}(\mathbf{x})^\top \nabla_{\mathbf{y}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) - \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} - \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})^\top \mathbf{J}(\mathbf{x})\dot{\mathbf{x}} \\
&= \mathbf{J}(\mathbf{x})^\top (\mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})\dot{\mathbf{x}} + \overset{\times}{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})\mathbf{J}(\mathbf{x})\dot{\mathbf{x}} - \frac{1}{2}\nabla_{\mathbf{y}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}})) - \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})^\top \mathbf{J}(\mathbf{x})\dot{\mathbf{x}}
\end{aligned}$$

In the second to the last equality above, we use $\partial_{\mathbf{x}} \dot{\mathbf{y}} = \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})$ and derive

$$\begin{aligned}
\frac{1}{2}\nabla_{\mathbf{x}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) &= \frac{1}{2}\mathbf{J}^\top \nabla_{\mathbf{y}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) + \frac{1}{2}\nabla_{\mathbf{x}}(\dot{\mathbf{y}})\nabla_{\dot{\mathbf{y}}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) \\
&= \frac{1}{2}\mathbf{J}^\top \nabla_{\mathbf{y}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) + \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} + \frac{1}{2}\dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \nabla_{\dot{\mathbf{y}}}(\mathbf{z}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{z}})|_{\mathbf{z}=\dot{\mathbf{y}}} \\
&= \frac{1}{2}\mathbf{J}^\top \nabla_{\mathbf{y}}(\dot{\mathbf{y}}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}}) + \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{y}} + \dot{\mathbf{J}}(\mathbf{x},\dot{\mathbf{x}})^\top \boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})^\top \mathbf{J}(\mathbf{x})\dot{\mathbf{x}}
\end{aligned}$$

as $\frac{1}{2}\partial_{\dot{\mathbf{y}}}(\mathbf{z}^\top \mathbf{N}(\mathbf{y},\dot{\mathbf{y}})\dot{\mathbf{z}})|_{\mathbf{z}=\dot{\mathbf{y}}} = \frac{1}{2}\dot{\mathbf{y}}^\top \left(\sum_{i=1}^{n} \dot{y}_i \partial_{\dot{\mathbf{y}}} \mathbf{n}_i(\mathbf{y},\dot{\mathbf{y}})\right) = \dot{\mathbf{y}}^\top \boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y},\dot{\mathbf{y}})$, where $\mathbf{n}_i$ is the ith column of $\mathbf{N}$.

To further simplify the expression, we note that by $\partial_{\mathbf{x}}\dot{\mathbf{y}} = \dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})$ we have

$$\begin{aligned}
\overset{\mathbf{x}}{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}} &= \sum_{i=1}^{n} \dot{y}_i \partial_{\mathbf{x}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{x}} \\
&= \sum_{i=1}^{n} \dot{y}_i (\partial_{\mathbf{y}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}})\mathbf{J}(\mathbf{x})\dot{\mathbf{x}} + \partial_{\dot{\mathbf{y}}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}})\partial_{\mathbf{x}}(\dot{\mathbf{y}})\dot{\mathbf{x}}) \\
&= \sum_{i=1}^{n} \dot{y}_i \partial_{\mathbf{y}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}} + \dot{y}_i \partial_{\dot{\mathbf{y}}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} \\
&= \left(\sum_{i=1}^{n} \dot{y}_i \partial_{\mathbf{y}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}}) \right) \dot{\mathbf{y}} + \left(\sum_{i=1}^{n} \dot{y}_i \partial_{\dot{\mathbf{y}}} \mathbf{n}_i(\mathbf{y}, \dot{\mathbf{y}}) \right) \dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} \\
&= \overset{\mathbf{y}}{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}} + 2\boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}}
\end{aligned}$$

Combining these two equalities, we can write

$$\begin{aligned}
\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{J}(\mathbf{x})^{\top} \Big(&\overset{\mathbf{y}}{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}} - \frac{1}{2}\nabla_{\mathbf{y}}(\dot{\mathbf{y}}^{\top}\mathbf{N}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}}) + (\mathbf{N}(\mathbf{y}, \dot{\mathbf{y}}) + 2\boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}}))\dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} \Big) \\
&- \dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})^{\top}\boldsymbol{\Xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})^{\top}\mathbf{J}(\mathbf{x})\dot{\mathbf{x}}
\end{aligned}$$

Substituting the definition of $\boldsymbol{\xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}}) = \overset{\mathbf{y}}{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}} - \frac{1}{2}\nabla_{\mathbf{y}}(\dot{\mathbf{y}}^{\top}\mathbf{N}(\mathbf{y}, \dot{\mathbf{y}})\dot{\mathbf{y}})$ proves the general statement.

In the special case, $\mathbf{M}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{M}(\mathbf{x})$ (which implies $\boldsymbol{\Xi}_{\mathbf{M}} = 0$),

$$\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{J}(\mathbf{x})^{\top} \Big(\boldsymbol{\xi}_{\mathbf{N}}(\mathbf{y}, \dot{\mathbf{y}}) + \mathbf{N}(\mathbf{y})\dot{\mathbf{J}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} \Big)$$

We show this expression is equal to $\boldsymbol{\eta}_{\mathbf{M};\mathcal{S}}$ regardless of the structure $\mathcal{S}$. This can be seen from the follows: If further $\mathbf{N}(\mathbf{y}) = \mathbf{L}(\mathbf{y})^{\top}\mathbf{C}(\mathbf{z})\mathbf{L}(\mathbf{y})$ and $\mathbf{M}$ is structured as $(\mathbf{LJ})^{\top}\mathbf{C}(\mathbf{LJ})$ from some Jacobian matrix $\mathbf{L}(\mathbf{y}) = \partial_{\mathbf{y}}\mathbf{z}$, we can write

$$\begin{aligned}
\boldsymbol{\eta}_{\mathbf{M};\mathcal{S}} &= \mathbf{J}^{\top}\mathbf{L}^{\top}(\boldsymbol{\xi}_{\bar{\mathbf{C}}} + \mathbf{C}\frac{d(\mathbf{LJ})}{dt}\dot{\mathbf{x}}) \\
&= \mathbf{J}^{\top}(\mathbf{L}^{\top}\boldsymbol{\xi}_{\bar{\mathbf{C}}} + \mathbf{L}^{\top}\mathbf{C}(\dot{\mathbf{L}}\mathbf{J} + \mathbf{L}\dot{\mathbf{J}})\dot{\mathbf{x}}) \\
&= \mathbf{J}^{\top} \Big(\mathbf{L}^{\top}(\boldsymbol{\xi}_{\bar{\mathbf{C}}} + \mathbf{C}\dot{\mathbf{L}}\dot{\mathbf{y}}) + \mathbf{L}^{\top}\mathbf{C}\mathbf{L}\dot{\mathbf{J}}\dot{\mathbf{x}} \Big) \\
&= \mathbf{J}^{\top} \Big(\boldsymbol{\xi}_{\mathbf{N}} + \mathbf{N}\dot{\mathbf{J}}\dot{\mathbf{x}} \Big) = \boldsymbol{\xi}_{\mathbf{M}}
\end{aligned}$$

■

B.2 Proof of Proposition 1

Proposition 1. *For* $(\mathcal{C}, \mathbf{G}, \mathbf{B}, \Phi)_{\mathcal{S}}$, $\dot{\mathbf{V}}(\mathbf{q}, \dot{\mathbf{q}}) = -\dot{\mathbf{q}}^{\top}\mathbf{B}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}}$.

Proof of Proposition 1. Let $K(\mathbf{q},\dot{\mathbf{q}}) = \frac{1}{2}\dot{\mathbf{q}}^\top \mathbf{G}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}$. Its time derivative can be written as

$$\begin{aligned}
\frac{d}{dt}K(\mathbf{q},\dot{\mathbf{q}}) &= \dot{\mathbf{q}}^\top \left(\mathbf{G}(\mathbf{q},\dot{\mathbf{q}})\ddot{\mathbf{q}} + \frac{1}{2}(\frac{d}{dt}\mathbf{G}(\mathbf{q},\dot{\mathbf{q}}))\dot{\mathbf{q}}\right) \\
&= \dot{\mathbf{q}}^\top \left(\mathbf{G}(\mathbf{q},\dot{\mathbf{q}})\ddot{\mathbf{q}} + \frac{1}{2}\sum_{i=1}^{d}\dot{q}_i\frac{d}{dt}\mathbf{g}_i(\mathbf{q},\dot{\mathbf{q}})\right) \\
&= \dot{\mathbf{q}}^\top \left(\mathbf{G}(\mathbf{q},\dot{\mathbf{q}})\ddot{\mathbf{q}} + \frac{1}{2}\sum_{i=1}^{d}\dot{q}_i\partial_{\mathbf{q}}\mathbf{g}_i(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \frac{1}{2}\sum_{i=1}^{d}\dot{q}_i\partial_{\dot{\mathbf{q}}}\mathbf{g}_i(\mathbf{q},\dot{\mathbf{q}})\ddot{\mathbf{q}}\right) \\
&= \dot{\mathbf{q}}^\top \left((\mathbf{G}(\mathbf{q},\dot{\mathbf{q}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}}))\ddot{\mathbf{q}} + \frac{1}{2}\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}\right)
\end{aligned}$$

where we recall $\mathbf{G}$ is symmetric and $\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}}) \coloneqq [\partial_{\mathbf{q}}\mathbf{g}_i(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}]_{i=1}^{d}$. Therefore, by definition $(\mathbf{G}(\mathbf{q},\dot{\mathbf{q}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}}))\ddot{\mathbf{q}} = (-\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q},\dot{\mathbf{q}}) - \nabla_{\mathbf{q}}\Phi(\mathbf{q}) - \mathbf{B}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}(\mathbf{q},\dot{\mathbf{q}}))$, we can derive

$$\begin{aligned}
\frac{d}{dt}\mathbf{V}(\mathbf{q},\dot{\mathbf{q}}) &= \frac{d}{dt}K(\mathbf{q},\dot{\mathbf{q}}) + \dot{\mathbf{q}}^\top\nabla_{\mathbf{q}}\Phi(\mathbf{q}) \\
&= \dot{\mathbf{q}}^\top \left(-\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q},\dot{\mathbf{q}}) - \nabla_{\mathbf{q}}\Phi(\mathbf{q}) - \mathbf{B}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{x}} + \frac{1}{2}\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \nabla_{\mathbf{q}}\Phi(\mathbf{q})\right) \\
&= -\dot{\mathbf{q}}^\top\mathbf{B}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \dot{\mathbf{q}}^\top \left(-\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q},\dot{\mathbf{q}}) + \frac{1}{2}\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}\right)
\end{aligned}$$

To finish the proof, we use two lemmas below.

Lemma 4. $\frac{1}{2}\dot{\mathbf{q}}^\top\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} = \dot{\mathbf{q}}^\top\boldsymbol{\xi}_{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})$.

Proof of Lemma 4. This can be shown by definition:

$$\begin{aligned}
\dot{\mathbf{q}}^\top\boldsymbol{\xi}_{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}}) &= \dot{\mathbf{q}}^\top \left(\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} - \frac{1}{2}\nabla_{\mathbf{q}}(\dot{\mathbf{q}}^\top\mathbf{G}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}})\right) \\
&= \sum_{k=1}^{d}\dot{q}_k\left(\sum_{i,j=1}^{d}\dot{q}_i\dot{q}_j\partial_{q_j}G_{k,i} - \frac{1}{2}\sum_{i,j=1}^{d}\dot{q}_i\dot{q}_j\partial_{q_k}G_{i,j}\right) \\
&= \sum_{i,j,k=1}^{d}\dot{q}_i\dot{q}_j\dot{q}_k\partial_{q_j}G_{k,i} - \frac{1}{2}\sum_{i,j,k=1}^{d}\dot{q}_i\dot{q}_j\dot{q}_k\partial_{q_k}G_{i,j} \\
&= \sum_{i,j,k=1}^{d}\dot{q}_i\dot{q}_j\dot{q}_k\partial_{q_k}G_{j,i} - \frac{1}{2}\sum_{i,j,k=1}^{d}\dot{q}_i\dot{q}_j\dot{q}_k\partial_{q_k}G_{i,j} \\
&= \frac{1}{2}\sum_{i,j,k=1}^{d}\dot{q}_i\dot{q}_j\dot{q}_k\partial_{q_j}G_{k,i} = \frac{1}{2}\dot{\mathbf{q}}^\top\overset{\mathbf{q}}{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}
\end{aligned}$$

where for the second to the last equality we use the symmetry $G_{i,j} = G_{j,i}$. ■

Using Lemma 4, we can show another equality.

Lemma 5. *For all structure* $\mathcal{S}$, $\dot{\mathbf{q}}^\top \left(-\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q},\dot{\mathbf{q}}) + \frac{1}{2}\overset{\mathbf{q}}{\dot{\mathbf{G}}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}\right) = 0$

Proof of Lemma 5. This can be seen from Lemma 3. Suppose $\mathcal{S}$ factorizes $\mathbf{G}(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{J}(\mathbf{q})^\top \mathbf{H}(\mathbf{x},\dot{\mathbf{x}})\mathbf{J}(\mathbf{q})$ where $\mathbf{J}(\mathbf{q}) = \partial_{\mathbf{q}}\mathbf{x}$. By Lemma 3, we know

$$\boldsymbol{\xi}_{\mathbf{G}} = \mathbf{J}^\top \left(\boldsymbol{\xi}_{\mathbf{H}} + (\mathbf{H} + 2\boldsymbol{\Xi}_{\mathbf{H}})\dot{\mathbf{J}}\dot{\mathbf{x}}\right) - \dot{\mathbf{J}}^\top \boldsymbol{\Xi}_{\mathbf{H}}^\top \mathbf{J}\dot{\mathbf{x}}$$

On the other hand, by definition, we have $\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}} \coloneqq \mathbf{J}^\top(\boldsymbol{\xi}_{\mathbf{H}} + (\mathbf{H} + \boldsymbol{\Xi}_{\mathbf{H}})\dot{\mathbf{J}}\dot{\mathbf{x}})$. Therefore, by comparing the two, we can derive,

$$\dot{\mathbf{q}}^\top \boldsymbol{\xi}_{\mathbf{G}} = \dot{\mathbf{q}}^\top \left(\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}} + \mathbf{J}^\top \boldsymbol{\Xi}_{\mathbf{H}} \dot{\mathbf{J}}\dot{\mathbf{q}} - \dot{\mathbf{J}}^\top \boldsymbol{\Xi}_{\mathbf{H}}^\top \mathbf{J}\dot{\mathbf{q}}\right) = \dot{\mathbf{q}}^\top \boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}$$

Combing the above equality and Lemma 4 proves the equality. ■

Finally, we use Lemma 5 and the previous result and conclude

$$\frac{d}{dt}\mathbf{V}(\mathbf{q},\dot{\mathbf{q}}) = -\dot{\mathbf{q}}^\top \mathbf{B}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \dot{\mathbf{q}}^\top \left(-\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q},\dot{\mathbf{q}}) + \frac{1}{2}\overset{\mathbf{q}}{\dot{\mathbf{G}}}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}\right) = -\dot{\mathbf{q}}^\top \mathbf{B}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}$$ ■

B.3 Proof of Theorem 2

Theorem 2. *Suppose every leaf node is a GDS with a metric matrix in the form* $\mathbf{R}(\mathbf{x}) + \mathbf{L}(\mathbf{x})^\top \mathbf{D}(\mathbf{x},\dot{\mathbf{x}})\mathbf{L}(\mathbf{x})$ *for differentiable functions* $\mathbf{R}$, $\mathbf{L}$, *and* $\mathbf{D}$ *satisfying* $\mathbf{R}(\mathbf{x}) \succeq 0$, $\mathbf{D}(\mathbf{x},\dot{\mathbf{x}}) = \operatorname{diag}((d_i(\mathbf{x},\dot{y}_i))_{i=1}^n) \succeq 0$, *and* $\dot{y}_i \partial_{\dot{y}_i} d_i(\mathbf{x},\dot{y}_i) \geq 0$, *where* $\mathbf{x}$ *is the coordinate of the leaf-node manifold and* $\dot{\mathbf{y}} = \mathbf{L}\dot{\mathbf{x}} \in \mathbb{R}^n$. *It holds* $\boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q},\dot{\mathbf{q}}) \succeq 0$. *If further* $\mathbf{G}(\mathbf{q},\dot{\mathbf{q}}), \mathbf{B}(\mathbf{q},\dot{\mathbf{q}}) \succ 0$, *then* $\mathbf{M} \in \mathbb{R}_{++}^{d\times d}$, *and the global RMP generated by RMPflow converges to the forward invariant set* $\mathcal{C}_\infty$ *in Corollary 2.*

Proof. Let $\mathbf{A}(\mathbf{x},\dot{\mathbf{x}}) = \mathbf{R}(\mathbf{x}) + \mathbf{L}(\mathbf{x})^\top \mathbf{D}(\mathbf{x},\dot{\mathbf{x}})\mathbf{L}(\mathbf{x})$. The proof of the theorem is straightforward, if we show that $\boldsymbol{\Xi}_{\mathbf{A}}(\mathbf{x},\dot{\mathbf{x}}) \succeq 0$. To see this, suppose $\mathbf{L} = \mathbb{R}^{n\times m}$. Let $\boldsymbol{\omega}_j^\top$ be the jth row $\mathbf{L}$, respectively. By definition of $\boldsymbol{\Xi}_{\mathbf{A}}(\mathbf{x},\dot{\mathbf{x}})$ we can write

$$\begin{aligned}
\boldsymbol{\Xi}_{\mathbf{A}}(\mathbf{x},\dot{\mathbf{x}}) &= \frac{1}{2}\sum_{i=1}^m \dot{x}_i \partial_{\dot{\mathbf{x}}} \mathbf{a}_i(\mathbf{x},\dot{\mathbf{x}}) \\
&= \frac{1}{2}\mathbf{L}(\mathbf{x})^\top \sum_{i=1}^m \dot{x}_i \partial_{\dot{\mathbf{x}}} (\mathbf{D}(\mathbf{x},\dot{\mathbf{x}})\mathbf{l}_i(\mathbf{x})) \\
&= \frac{1}{2}\mathbf{L}(\mathbf{x})^\top \sum_{i=1}^m \sum_{j=1}^n \dot{x}_i \partial_{\dot{\mathbf{x}}} (d_j(\mathbf{x},\dot{y}_j) L_{ji}(\mathbf{x})\mathbf{e}_j) \\
&= \frac{1}{2}\mathbf{L}(\mathbf{x})^\top \sum_{j=1}^n \left(\sum_{i=1}^m L_{ji}(\mathbf{x})\dot{x}_i\right) \partial_{\dot{y}_j} d_j(\mathbf{x},\dot{y}_j)\mathbf{e}_j \boldsymbol{\omega}_j^\top \\
&= \frac{1}{2}\mathbf{L}(\mathbf{x})^\top \sum_{j=1}^n \dot{y}_j \partial_{\dot{y}_j} d_j(\mathbf{x},\dot{y}_j)\mathbf{e}_j \boldsymbol{\omega}_j^\top \\
&= \mathbf{L}(\mathbf{x})^\top \boldsymbol{\Xi}_{\mathbf{D}}(\mathbf{x},\dot{\mathbf{x}})\mathbf{L}(\mathbf{x})
\end{aligned}$$

where $\mathbf{e}_j$ the jth canonical basis and $\mathbf{\Xi_D}(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}\mathrm{diag}((\partial_{\dot{y}_i} d_i(\mathbf{x}, \dot{y}_i))_{i=1}^n)$. Therefore, under the assumption that $\partial_{\dot{y}_i} d_i(\mathbf{x}, \dot{y}_i) \geq 0$, $\mathbf{\Xi_A}(\mathbf{x}, \dot{\mathbf{x}}) \succeq 0$. This further implies $\mathbf{\Xi_G}(\mathbf{q}, \dot{\mathbf{q}}) \succeq 0$ by Theorem 1.

The stability of the entire system follows naturally from the rule of `pullback`, which ensures that $\mathbf{M}_r(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{M}(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{G}(\mathbf{q}, \dot{\mathbf{q}}) + \mathbf{\Xi_G}(\mathbf{q}, \dot{\mathbf{q}}) \succ 0$ given that the leaf-node condition is satisfied. Consequently, the condition in Corollary 2 holds and the convergence to $\mathcal{C}_\infty$ is guaranteed. ■

B.4 Notation for Coordinate-Free Analysis

We introduce some extra notations for the coordinate-free analysis. Let $p_{T\mathcal{C}} : T\mathcal{C} \to \mathcal{C}$ be the bundle projection. Suppose $(U, (\mathbf{q}, \mathbf{v}))$ is a (local) chart on $T\mathcal{C}$. Let $\{\frac{\partial}{\partial q_i}, \frac{\partial}{\partial v_i}\}_{i=1}^d$ and $\{\mathrm{d}q^i, \mathrm{d}v^i\}_{i=1}^d$ denote the induced frame field and coframe field on $T\mathcal{C}$. For $s \in U$, we write s in coordinate as $(\mathbf{q}(q), \mathbf{v}(s))$, if $\sum_{i=1}^d v_i(s)\frac{\partial}{\partial q_i}|_q \in T_q\mathcal{C}$, where $q = p_{T\mathcal{C}}(s) \in \mathcal{C}$. With abuse of notation, we also write $s = (\mathbf{q}, \mathbf{v})$ for short unless clarity is lost. Similarly, a chart $(\tilde{U}, (\mathbf{q}, \mathbf{v}, \mathbf{u}, \mathbf{a}))$ can naturally be constructed on the double tangent bundle $TT\mathcal{C}$, where $\tilde{U} = p_{TT\mathcal{C}}^{-1}(U)$ and $p_{TT\mathcal{C}} : TT\mathcal{C} \to T\mathcal{C}$ is the bundle projection: we write $h = (\mathbf{q}, \mathbf{v}, \mathbf{u}, \mathbf{a}) \in TT\mathcal{C}$ if $\sum_{i=1}^d u_i(h)\frac{\partial}{\partial q_i}|_s + a_i(h)\frac{\partial}{\partial v_i}|_s \in T_sT\mathcal{C}$, where $s = p_{TT\mathcal{C}}(h)$. Under these notations, for a curve $q(t)$ on $\mathcal{C}$, we can write $\ddot{q}(t) \in TT\mathcal{C}$ in coordinate as $(\mathbf{q}(t), \dot{\mathbf{q}}(t), \dot{\mathbf{q}}(t), \ddot{\mathbf{q}}(t))$. Finally, given Christoffel symbols $\Gamma_{i,j}^k$, an affine connection ∇ on $TT\mathcal{C}$ is defined via $\nabla_{\frac{\partial}{\partial s_i}}\frac{\partial}{\partial s_j} = \sum_{k=1}^{2d}\Gamma_{i,j}^k\frac{\partial}{\partial s_k}$, where $\frac{\partial}{\partial s_i} \coloneqq \frac{\partial}{\partial q_i}$ and $\frac{\partial}{\partial s_{i+d}} \coloneqq \frac{\partial}{\partial v_i}$ for $i = 1, \ldots, d$.

B.5 Proof of Theorem 3

Theorem 3. *Let G be a Riemannian metric on $T\mathcal{C}$ such that, for $s = (q, v) \in T\mathcal{C}$, $G(s) = G_{ij}^v(s)dq^i \otimes dq^j + G_{ij}^a dv^i \otimes dv^j$, where $G_{ij}^v(s)$ and G_{ij}^a are symmetric and positive-definite, and $G_{ij}^v(\cdot)$ is differentiable. Then there is a unique affine connection ${}^G\nabla$ that is compatible with G and satisfies, $\Gamma_{i,j}^k = \Gamma_{ji}^k$, $\Gamma_{i,j+d}^k = 0$, and $\Gamma_{i+d,j+d}^k = \Gamma_{j+d,i+d}^k$, for $i, j = 1, \ldots, d$ and $k = 1, \ldots, 2d$. In coordinates, if $G_{ij}^v(\dot{q})$ is identified as $\mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})$, then $\mathrm{pr}_3({}^G\nabla_{\ddot{q}}\ddot{q})$ can be written as $\mathbf{a_G} \coloneqq \ddot{\mathbf{q}} + \mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})^{-1}(\boldsymbol{\xi}_\mathbf{G}(\mathbf{q}, \dot{\mathbf{q}}) + \mathbf{\Xi_G}(\mathbf{q}, \dot{\mathbf{q}})\ddot{\mathbf{q}})$, where $\mathrm{pr}_3 : (\mathbf{q}, \mathbf{v}, \mathbf{u}, \mathbf{a}) \mapsto \mathbf{u}$ is a projection.*

Proof of Theorem 3. We first show ${}^G\nabla$ is unique, if it exists. That is, there is at most one affine connection that is compatible with the given the Riemannian metric G and satisfies for $i, j = 1, \ldots, d$ and $k = 1, \ldots, 2d$

$$\Gamma_{i,j}^k = \Gamma_{ji}^k, \qquad \Gamma_{i,j+d}^k = 0, \qquad \Gamma_{i+d,j+d}^k = \Gamma_{j+d,i+d}^k,$$

Importantly, we note that this definition is coordinate-free, independent of the choice of chart on $\mathcal{C}$.

The uniqueness is easy to see. As G is non-degenerate by definition, we recall there is an unique Levi-Civita connection, which is compatible with G and

satisfies the symmetric condition

$$\Gamma^k_{i,j} = \Gamma^k_{j,i}, \qquad \text{for } i,j = 1,\dots,2d$$

Comparing our asymmetric condition and the symmetric condition of the Levi-Civita connection, we see that number of the linearly independent constraints are the same; therefore if there is a solution to the required asymmetric affine connection, then it is unique.

Next we show such a solution exists. We consider the candidate Christoffel symbols below and show that they satisfy the requirements: Consider an *arbitrary* choice of chart on $\mathcal{C}$. For $i,j,k = 1,\dots,d$,

$$\begin{aligned}
\Gamma^k_{i,j} &= \frac{1}{2}\sum_{l=1}^{d} G^{v\sharp}_{k,l}(\partial_{q_j} G^v_{l,i} + \partial_{q_i} G^v_{l,j} - \partial_{q_l} G^v_{i,j}) \\
\Gamma^k_{i,j+d} = 0, \quad \Gamma^k_{i+d,j} &= \frac{1}{2}\sum_{l=1}^{d} G^{v\sharp}_{k,l}(\partial_{v_i} G^v_{l,j}), \quad \Gamma^k_{i+d,j+d} = 0 \\
\Gamma^{k+d}_{i,j} = 0, \quad \Gamma^{k+d}_{i,j+d} = 0, &\quad \Gamma^{k+d}_{i+d,j} = 0, \quad \Gamma^{k+d}_{i+d,j+d} = 0
\end{aligned}$$

where $G^{v\sharp}$ denotes the inverse of G^v, i.e. $\sum_{k=1}^{d} G^{v\sharp}_{i,k} G^v_{k,j} = \delta_{i,j}$. Note we choose not to adopt the Einstein summation notation, so the sparse pattern of the proposed Christoffel symbols are clear.

It is clear that the above candidate Christoffel symbols satisfies the asymmetric condition. Therefore, to show it is a solution, we only need to show such choice is compatible with G. Equivalently, it means for arbitrary smooth sections of $TT\mathcal{C}$, $X = \sum_{i=1}^{2d} X_i \frac{\partial}{\partial s_i}$, $Y = \sum_{i=1}^{2d} Y_i \frac{\partial}{\partial s_i}$, $Z = \sum_{i=1}^{2d} Z_i \frac{\partial}{\partial s_i}$, it holds[9]

$$^G\nabla_Z G(X,Y) = G(^G\nabla_Z X, Y) + G(X, ^G\nabla_Z Y) \tag{8}$$

To verify (8), we first write out $^G\nabla_Z X$ using the chosen Christoffel symbols:

$$\begin{aligned}
^G\nabla_Z X &= \sum_{k=1}^{2d} \left(^G\nabla_Z X_k + \sum_{i,j=1}^{2d} \Gamma^k_{ij} Z_i X_j \right) \frac{\partial}{\partial s_k} \\
&= \sum_{k=1}^{d} D_Z(X_k) \frac{\partial}{\partial q_k} + \sum_{k=1}^{d} D_Z(X_{k+d}) \frac{\partial}{\partial v_k} \\
&\quad + \frac{1}{2}\sum_{k,l=1}^{d} G^{v,kl} \left(\sum_{i,j,=1}^{d} (\partial_{q_j} G^v_{li} + \partial_{q_i} G^v_{lj} - \partial_{q_l} G^v_{ij}) Z_i X_j + (\partial_{v_i} G^v_{lj}) Z_{i+d} X_j \right) \frac{\partial}{\partial q_k}
\end{aligned} \tag{9}$$

[9] The section requirement on Z can be dropped.

where $D_Z(\cdot)$ denotes the derivation with respect to Z. The above implies

$$G({}^G\nabla_Z X, Y) = \sum_{j,k=1}^{d} G^v_{ki} Y_k D_Z(X_i) + \sum_{j,k=1}^{d} G^a_{kj} Y_{k+d} D_Z(X_{j+d}) + \frac{1}{2}\left(\sum_{i,j,k=1}^{d} (\partial_{q_j} G^v_{ki} + \partial_{q_i} G^v_{kj} - \partial_{q_k} G^v_{ij}) Z_i X_j Y_k + (\partial_{v_i} G^v_{kj}) Z_{i+d} X_j Y_k \right)$$

Similarly, we can derive $G(X, {}^G\nabla_Z Y)$. Using the symmetry $G^v_{ij} = G^v_{ji}$, we can combine the previous results together and write

$$\begin{aligned}
&G({}^G\nabla_Z X, Y) + G(X, {}^G\nabla_Z Y) \\
&= \sum_{j,k=1}^{d} G^v_{ki} Y_k D_Z(X_i) + \sum_{j,k=1}^{d} G^a_{kj} Y_{k+d} D_Z(X_{j+d}) + \sum_{j,k=1}^{d} G^v_{ki} X_k D_Z(Y_i) + \sum_{j,k=1}^{d} G^a_{kj} X_{k+d} D_Z(Y_{j+d}) \\
&\quad + \frac{1}{2}\left(\sum_{i,j,k=1}^{d} (\partial_{q_j} G^v_{ki} + \partial_{q_i} G^v_{kj} - \partial_{q_k} G^v_{ij}) Z_i X_j Y_k + (\partial_{v_i} G^v_{kj}) Z_{i+d} X_j Y_k \right) \\
&\quad + \frac{1}{2}\left(\sum_{i,j,k=1}^{d} (\partial_{q_j} G^v_{ki} + \partial_{q_i} G^v_{kj} - \partial_{q_k} G^v_{ij}) Z_i Y_j X_k + (\partial_{v_i} G^v_{kj}) Z_{i+d} Y_j X_k \right) \\
&= \sum_{i,j=1}^{d} G^v_{ij} D_Z(X_i) Y_j + G^v_{ij} X_i D_Z(Y_j) + \sum_{i,j=1}^{d} G^a_{ij} D_Z(X_{d+i}) Y_{d+j} + G^a_{ij} X_{d+i} D_Z(Y_{d+j}) \\
&\quad + \sum_{i,j,k=1}^{d} X_i Y_j Z_k \partial_{q_k} G^v_{ij} + X_i Y_j Z_{k+d} \partial_{v_k} G^v_{ij} \\
&= \sum_{i,j=1}^{d} D_Z(G^v_{ij}) X_i Y_j + G^v_{ij} D_Z(X_i) Y_j + G^v_{ij} X_i D_Z(Y_j) + \sum_{i,j=1}^{d} G^a_{ij} D_Z(X_{d+i}) Y_{d+j} + G^a_{ij} X_{d+i} D_Z(Y_{d+j}) \\
&= {}^G\nabla_Z \left(\sum_{i,j=1}^{d} G^v_{ij} X_i Y_j + \sum_{i,j=1}^{d} G^a_{ij} X_{d+i} Y_{d+j} \right) = {}^G\nabla_Z G(X, Y)
\end{aligned}$$

Therefore ${}^G\nabla$ is compatible with G.

So far we have proved the first statement of Theorem 3 that ${}^G\nabla$ is the unique solution that is compatible with G and satisfies the asymmetric condition. Below we show the expression of $\mathrm{pr}_3({}^G\nabla_{\dot{q}}\ddot{q})$, where we recall $\ddot{q}(t)$ is a curve in $TT\mathcal{C}$. We use (9). By definition of pr_3 it extracts the parts on $\{\frac{\partial}{\partial q_i}\}_{i=1}^d$. Therefore,

suppose we choose some chart on $\mathcal{C}$ of interest and we can write $\mathrm{pr}_3({}^G\nabla_{\ddot{q}}\ddot{q})$ as

$$\begin{aligned}\mathrm{pr}_3({}^G\nabla_{\ddot{q}}\ddot{q}) &= \sum_k^d \left(D_Z(X_k) + \sum_{l=1}^d \frac{1}{2} G^{v,kl} \sum_{i,j,=1}^d (\partial_{q_j} G^v_{li} + \partial_{q_i} G^v_{lj} - \partial_{q_l} G^v_{ij}) Z_i X_j + (\partial_{v_i} G^v_{lj}) Z_{i+d} X_j \right) \frac{\partial}{\partial q_k} \\ &= \sum_k^d \left(\ddot{q}_k + \sum_{l=1}^d \frac{1}{2} G^{v,kl} \sum_{i,j,=1}^d (\partial_{q_j} G^v_{li} + \partial_{q_i} G^v_{lj} - \partial_{q_l} G^v_{ij}) \dot{q}_i \dot{q}_j + (\partial_{v_i} G^v_{lj}) \ddot{q}_i \dot{q}_j \right) \frac{\partial}{\partial q_k} \\ &= \sum_k^d a_{\mathbf{G};k} \frac{\partial}{\partial q_k}\end{aligned}$$

where $a_{\mathbf{G};k}$ is the kth element of $\mathbf{a}_{\mathbf{G}} \coloneqq \ddot{\mathbf{q}} + \mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})^{-1}(\boldsymbol{\xi}_{\mathbf{G}}(\mathbf{q}, \dot{\mathbf{q}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q}, \dot{\mathbf{q}})\ddot{\mathbf{q}})$. ■

B.6 Proof of Theorem 4

Theorem 4. *Suppose $\mathcal{C}$ is related to K leaf-node task spaces by maps $\{\psi_i : \mathcal{C} \to \mathcal{T}_i\}_{i=1}^K$ and the ith task space $\mathcal{T}_i$ has an affine connection ${}^{G_i}\nabla$ on $T\mathcal{T}_i$, as defined in Theorem 3, and a covector function F_i defined by some potential and damping as described above. Let ${}^G\bar{\nabla} = \sum_{i=1}^K T\psi_i^{*G_i}\nabla$ be the pullback connection, $G = \sum_{i=1}^K T\psi_i^* G_i$ be the pullback metric, and $F = \sum_{i=1}^K T\psi_i^* F_i$ be the pullback covector, where $T\psi_i^* : T^*T\mathcal{T}_i \to T^*T\mathcal{C}$. Then ${}^G\bar{\nabla}$ is compatible with G, and $\mathrm{pr}_3({}^G\bar{\nabla}_{\ddot{q}}\ddot{q}) = (\mathrm{pr}_3 \circ G^\sharp \circ F)(s)$ can be written as $\ddot{\mathbf{q}} + \mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})^{-1}(\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q}, \dot{\mathbf{q}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q}, \dot{\mathbf{q}})\ddot{\mathbf{q}}) = -\mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})^{-1}(\nabla_{\mathbf{q}}\Phi(\mathbf{q}) + \mathbf{B}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}})$. In particular, if G is velocity-independent, then ${}^G\bar{\nabla} = {}^G\nabla$.*

Proof of Theorem 4. Let $\mathcal{T} = \mathcal{T}_1 \times \cdots \times \mathcal{T}_K$ and $\tilde{G}$ be the induced metric on $T\mathcal{T}$ by $\{G_i\}_{i=1}^K$. In addition, let $\psi : \mathcal{C} \to \mathcal{T}$ be the equivalent expression of $\{\psi_i\}$. Again we focus on the tangent bundle not the base manifold. Recall the definition of a pullback connection[10] $T\psi^{*\tilde{G}}\nabla$ is

$$T\psi_*(T\psi^{*\tilde{G}}\nabla_X Y) = \mathrm{pr}^{\tilde{G}}_{T\psi_*}\left({}^{\tilde{G}}\nabla_{T\psi_* X} T\psi_* Y\right) \tag{10}$$

for all sections X and Y on $T\mathcal{C}$, where $\mathrm{pr}^{\tilde{G}}_{T\psi_*}$ is the projection onto the distribution spanned by $T\psi_*$ with respect to $\tilde{G}$, i.e. $\tilde{G}(T\psi_* X, \mathrm{pr}^{\tilde{G}}_{T\psi_*}(Z)) = \tilde{G}(T\psi_* X, Z)$ for all $X \in T\mathcal{C}$ and $Z \in T\mathcal{T}$. Note by the construction of the product manifold $\mathcal{T}$, $T\psi^{*\tilde{G}}\nabla = \sum_{i=1}^K T\psi_i^{*G_i}\nabla$.

Compatibility We show that $T\psi^{*\tilde{G}}\nabla$ is compatible with the pullback metric G. Let X, Y, Z be arbitrary sections on $T\mathcal{C}$ and recall the definition of the pullback metric

$$G(X, Y) = T\psi^*\tilde{G}(X, Y) = \tilde{G}(T\psi_* X, T\psi_* Y)$$

[10] We note the distinction between $\psi^* : T^*\mathcal{T} \to T^*\mathcal{C}$ and $T\psi^* : T^*T\mathcal{T} \to T^*T\mathcal{C}$.

To show that $T\psi^{*\tilde{G}}\nabla$ is compatible, we derive an expression of $G(T\psi^{*\tilde{G}}\nabla_Z X, Y)$:

$$\begin{aligned} G(T\psi^{*\tilde{G}}\nabla_Z X, Y) &= T\psi^*\tilde{G}(T\psi^{*\tilde{G}}\nabla_Z X, Y) \\ &= \tilde{G}(T\psi_*(T\psi^{*\tilde{G}}\nabla_Z X), T\psi_* Y) \\ &= \tilde{G}\left(\mathrm{pr}^{\tilde{G}}_{T\psi_*}\left({}^{\tilde{G}}\nabla_{T\psi_* Z} T\psi_* X\right), T\psi_* Y\right) \\ &= \tilde{G}\left({}^{\tilde{G}}\nabla_{T\psi_* Z} T\psi_* X, T\psi_* Y\right) \end{aligned}$$

where we use (10) and the definition of projection. Using the above equation, we can see the compatibility easily:

$$\begin{aligned} &G(T\psi^{*\tilde{G}}\nabla_Z X, Y) + G(X, T\psi^{*\tilde{G}}\nabla_Z Y) \\ &= \tilde{G}\left({}^{\tilde{G}}\nabla_{T\psi_* Z} T\psi_* X, T\psi_* Y\right) + \tilde{G}\left(T\psi_* X, {}^{\tilde{G}}\nabla_{T\psi_* Z} T\psi_* Y\right) \\ &= {}^{\tilde{G}}\nabla_{T\psi_* Z}\tilde{G}(T\psi_* X, T\psi_* Y) \\ &= \psi^{*G}\nabla_Z \tilde{G}(T\psi_* X, T\psi_* Y) \\ &= \psi^{*G}\nabla_Z G(X, Y) \end{aligned}$$

Coordinate expression The coordinate expression of the pullback metric can be derived by its definition in (10), and the expression for the pullback covector is standard. For the pullback connection, similar to the proof of Theorem 3, we can show that $\mathrm{pr}_3({}^G\bar{\nabla}_{\ddot{q}}\ddot{q})$ can be written as $\ddot{\mathbf{q}} + \mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})^{-1}(\boldsymbol{\eta}_{\mathbf{G};\mathcal{S}}(\mathbf{q}, \dot{\mathbf{q}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q}, \dot{\mathbf{q}})\ddot{\mathbf{q}})$. In other words, the structured GDS equations are the coordinate expression of the pullback connection $T\psi^{*\tilde{G}}\nabla$, where the structure structure $\mathcal{S}$ is induced through the recursive application of `pullback` in RMPflow. Note that this is in general different from the connection of the pullback metric ${}^G\nabla$, which by Theorem 3 instead defines the unstructured GDS equation $\ddot{\mathbf{q}} + \mathbf{G}(\mathbf{q}, \dot{\mathbf{q}})^{-1}(\boldsymbol{\xi}_{\mathbf{G}}(\mathbf{q}, \dot{\mathbf{q}}) + \boldsymbol{\Xi}_{\mathbf{G}}(\mathbf{q}, \dot{\mathbf{q}})\ddot{\mathbf{q}})$.

Commutability However, in the special case when G is velocity-independent, we show that they are equivalent. That is, the pullback connection $T\psi^{*\tilde{G}}\nabla$ is equal to the connection of the pullback matrix ${}^G\nabla$. This property is early shown in Theorem 1, which shows that in the velocity-independent case there is no need to distinguish structures. To prove this, we first note that ${}^G\nabla$ becomes symmetric as G is velocity-independent. As it is also compatible with G, we know that ${}^G\nabla$ is the Levi-Civita connection with respect to G. (Recall G is the Riemannian metric on the tangent bundle.) On the other hand, knowing that $T\psi^{*\tilde{G}}\nabla$ is compatible, to show that ${}^G\nabla = T\psi^{*\tilde{G}}\nabla$ we only need to check if $T\psi^{*\tilde{G}}\nabla$ is symmetric. Without further details, we note this is implied by the proof of Theorem 1. Therefore, we have $T\psi^{*\tilde{G}}\nabla =^G \nabla$. ■

C Relationship between RMPflow and Recursive Newton-Euler Algorithms

The policy generation procedure of RMPflow is closely related to the algorithms [17] for computing forward dynamics (i.e. computing accelerations given forces) based on recursive Newton-Euler algorithm. In a summary, these algorithms computes the forward dynamics in following steps:

1. It propagates positions and velocities from the base to the end-effector.
2. It computes the Coriollis force by backward propagating the inverse dynamics of each link under the condition that the acceleration is zero.
3. It computes the (full/upper-triangular/lower-triangular) joint inertia matrix.
4. It solves a linear system of equations to obtain the joint acceleration.

In [17], they assume a recursive Newton-Euler algorithm (RNE) for inverse dynamics is given, and realize Step 1 and Step 2 above by calling the RNE subroutine. The computation of Step 3 depends on which part of the inertia matrix is computed. In particular, their Method 3 (also called the Composite-Rigid-Body Algorithm in [33, Chapter 6]) computes the upper triangle part of the inertia matrix by a backward propagation from the end-effector to the base.

RMPflow can also be used to compute forward dynamics, when we set the leaf-node GDS as the constant inertia system on the body frame of each link and we set the transformation in the RMP-tree as the change of coordinates across of robot links. This works because we show GDSs cover SMSs as a special case, and at root node the effective dynamics is the pullback GDS, which in this case is the effective robot dynamics defined by the inertia matrix of each link.

We can use this special case to compare RMPflow with the above procedure. We see that the forward pass of RMPflow is equivalent to Step 1, and the backward pass of RMPflow is equivalent of Step 2 and Step 3, and the final `resolve` operation is equivalent to Step 4.

Despite similarity, the main difference is that RMPflow computes the force and the inertia matrix in a *single* backward pass to exploit shared computations. This change is important, especially, the number of subtasks are large, e.g., in avoiding multiples obstacles. In addition, the design of RMPflow generalizes these classical computational procedures (e.g. designed only for rigid bodies, rotational/prismatic joints) to handle abstract and even non-Euclidean task spaces that have velocity-dependent metrics/inertias. This extension provides a unified framework of different algorithms and results in an expressive class of motion policies.

D Designing Reactive Motion Policies for Manipulation

In this section, we give some details on the RMPs examples discussed in Section 3.6, which are also used in our manipulation system in the full system experiments. We show that these commonly used motion policies are essentially

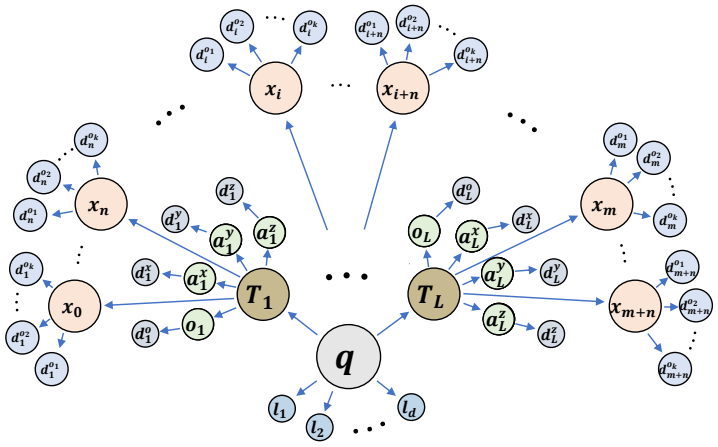

Fig. 6: This figure depicts the tree of task maps used in the experiments. See Section D.1 for details.

GDSs with respect to some metric and potential function. To convert a differential equation back to its GDS formulation, we need to address the question of integrability of a vector field. This is done by showing that a hand-designed vector field is the negative gradient of some potential function. It is useful in these cases to remember that the necessary and sufficient condition on the integrability of a smooth vector field is that its Jacobian is symmetric.

D.1 Task map and its Tree Structure

Figure 6 depicts the tree of task maps used in the full-robot experiments. The chosen structure emphasizes potential for parallelization over fully exploiting the recursive nature of the kinematic chain, treating each link frame as just one forward kinematic map step from the configuration space.[11] The configuration space $\mathbf{q}$ is linked to L link frames $\mathbf{T}_1, \ldots, \mathbf{T}_L$ through the robot's forward kinematics. Each frame has 4 frame element spaces: the origin o_i and each of the axes $\mathbf{a}_i^x, \mathbf{a}_i^y, \mathbf{a}_i^z$, with corresponding distance spaces to targets $d_i^o, d_i^x, d_i^y, d_i^z$ (if they are active). Additionally, there are a number of obstacle control points $\mathbf{x}_j$ distributed across each of the links, each with k associated distance spaces $d_j^{o_1}, \ldots, d_j^{o_k}$, one for each obstacle $o_1, \ldots, o_k$. Finally, for each dimension of the configuration space there's an associated joint limit space $l_1, \ldots, l_d$.

D.2 Example: 1D Velocity-Dependent Metrics

We start with an analysis of a simple 1-dimensional GDS with a velocity-dependent metric to provide some intuition about the curvature terms ξ and Ξ. This example will be used for constructing collision controllers later.

[11] We could possibly have saved some computation by defining the forward kinematic maps recursively as $(\mathbf{T}_{i+1}, \mathbf{q}_{i+1}, \ldots, \mathbf{q}_d) = \psi_i(\mathbf{T}_i, \mathbf{q}_i, \ldots, \mathbf{q}_d)$.

Let $z = d(\mathbf{x}) \in \mathbb{R}$ be a 1D task space; for instance, d might be a distance function with $\mathbf{x} \in \mathbb{R}^3$. Let $g(z, \dot{z})$ denote a velocity dependent metric and let $\Phi(z)$ be a potential function. This choice defines a total energy (i.e. the Lyapunov function) $V(z, \dot{z}) = \frac{1}{2} g(z, \dot{z}) \dot{z}^2 + \Phi(z)$. It defines a GDS with an equation of motion under external force $f_{\text{ext}} = -\partial_z \Phi - b(z, \dot{z})\dot{z}$ as

$$\ddot{z} = \frac{1}{g + \Xi} \left(-\partial_z \Phi - b(z, \dot{z})\dot{z} - \xi \right), \tag{11}$$

where $\Xi = \frac{1}{2} \dot{z} \frac{\partial g}{\partial \dot{z}}$, ξ is the curvature term (see below), and $b(z, \dot{z}) > 0$ is the damping coefficient.

Theorem 2 provides a sufficient condition for stability. In this example, it requires $\dot{z} \frac{\partial g}{\partial \dot{z}} \geq 0$. Suppose the metric decomposes as $g(z, \dot{z}) = w(z)u(\dot{z})$. The sufficient condition of Theorem 2 becomes

$$2\Xi = \dot{z} \frac{\partial g}{\partial \dot{z}} = w(z) \dot{z} \frac{du}{d\dot{z}} \geq 0, \tag{12}$$

In other words, u needs to change (as a function of $\dot{z}$) in the same direction as the velocity: u either increases in the positive direction when velocity is positive and increases in the negative direction when velocity is negative; or it can be zero.

Denoting $g_{\dot{z}}(z)$ as the corresponding function of $\dot{z}$ that arises by fixing the value of $\dot{z}$, we can write the curvature terms as

$$\xi = \left(\frac{d}{dt} g_{\dot{z}} \right) \dot{z} - \frac{d}{dz} \left(\frac{1}{2} g(z, \dot{z}) \dot{z}^2 \right) \tag{13}$$

$$= u(\dot{z}) \left(\frac{d}{dt} w(z) \right) \dot{z} - u(\dot{z}) \frac{d}{dz} \left(\frac{1}{2} w(z) \dot{z}^2 \right) \tag{14}$$

$$= u(\dot{z}) \left(\frac{dw}{dz} \dot{z}^2 - \frac{1}{2} \frac{dw}{dz} \dot{z}^2 \right) \tag{15}$$

$$= \frac{1}{2} u \frac{dw}{dz} \dot{z}^2. \tag{16}$$

Therefore $-\xi = -\frac{1}{2} u \frac{dw}{dz} \dot{z}^2$ is a force that always points along decreasing w.

D.3 Collision Avoidance Controllers

Here we derive a class of 1D collision controllers defined on the distance space, and show that the curvature terms in the pullback to $\mathbf{x} \in \mathbb{R}^3$ define nontrivial curving terms that induce the types of orbits that we see in Figure 3.

Let $s = d(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^3$ denote a distance function $d : \mathbb{R}^3 \to \mathbb{R}_+$. Let $g(s, \dot{s}) = w(s)u(\dot{s})$ denote a 1D separable velocity-dependent metric. $w(s)$ is defined as a non-increasing function in s, i.e. and $s_1 \leq s_2 \Rightarrow w(s_1) \geq w(s_2)$ and hence $\frac{dw}{ds} \leq 0$ for all $s \in \mathbb{R}_+$. Typically, $w(s) = 0$ for $s > r_w$ for some nominal radius of action $r_w > 0$. For instance, we might choose

$$w(s) = \frac{(r_w - s)_+^2}{s}, \tag{17}$$

where $(v)_+ = \max\{0, v\}$. For this function, $w(r_w) = 0$ and differentiating it shows that $\frac{dw}{ds} = 1 - \frac{r_w^2}{s^2} < 0$ for $s \in (0, r_w)$ and 0 for $s \geq r_w$. The equality only holds at $s = r_w$ (i.e. when w is tangent to the s axis at $s = r_w$). Its Hessian can be shown as $\frac{d^2w}{ds^2} = \frac{2r_w^2}{s^3} > 0$ for $s > 0$ (i.e. positive definite). Likewise we choose

$$u(\dot{s}) = \begin{cases} 1 - \exp\left(-\frac{\dot{s}^2}{2\sigma^2}\right), & \text{for } \dot{s} < 0 \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

It is straightforward to show that this choice satisfies the condition in (12), and $u(\dot{s}) \in [0, 1)$ with a smooth transition to 0 for $\dot{s} \geq 0$.

Suppose $\Phi : \mathbb{R}_+ \to \mathbb{R}$ is a potential function that is continuously differentiable. As discussed, the GDS would have an equation of motion in the form

$$\ddot{s}^d = -\frac{1}{m}\frac{d\Phi_0}{ds} - \frac{1}{m}\xi$$

where we recall $m(s, \dot{s}) = g(s, \dot{s}) + \Xi(s, \dot{s})$ and Ξ and ξ are the curvature terms. Particularly, for our chosen product metric $g(s, \dot{s}) = w(s)u(\dot{s})$, we can write $\xi(s, \dot{s}) = \frac{1}{2}u(\dot{s})\frac{dw}{ds}\dot{s}^2$, $\Xi(s, \dot{s}) = \frac{1}{2}w(s)\dot{s}\frac{du}{d\dot{s}}$ (cf. the previous section), and

$$\begin{aligned} m(s, \dot{s}) &= w(s)u(s) + \frac{1}{2}w(s)\dot{s}\frac{du}{d\dot{s}} \\ &= w(s)\left(u(s) + \frac{1}{2}\dot{s}\frac{du}{d\dot{s}}\right) =: w(s)\delta(s, \dot{s}) \end{aligned}$$

which also has a product structure.

This factorization shows that the equation of motion can also be written as

$$\begin{aligned} \ddot{s}^d &= -\frac{1}{\delta}\left(\frac{1}{w}\frac{d\Phi}{ds}\right) - \frac{1}{m}\xi \\ &= -\frac{1}{\delta}\left(\frac{d\widetilde{\Phi}}{ds}\right) - \frac{1}{w}\tilde{\xi} \end{aligned}$$

for some function $\widetilde{\Phi}$ and $\tilde{\xi} = \frac{1}{2}\frac{dw}{ds}\dot{s}^2$ is the curvature term if w is considered as a (velocity-independent) metric. This identification is possible because $\frac{1}{w}\frac{d\Phi}{ds}$ is continuous and s is one-dimensional. Conversely, we can start with designing a continuous vector field $\frac{d\widetilde{\Phi}}{ds}$, i.e. choosing a vector field $\frac{d\widetilde{\Phi}}{ds}$ such that $\ddot{s}^d = -\frac{1}{\delta}\left(\frac{d\widetilde{\Phi}}{ds}\right) - \frac{1}{m}\xi$ has the desired behavior. And the above identification shows that this equation of motion is a GDS.

To represent the above equation as an RMP, which is useful if we choose to design behavior by directly defining $\frac{d\widetilde{\Phi}}{ds}$, we can write it in the natural form as

$$\left[-w\frac{d\widetilde{\Phi}}{ds} - \xi, m\right]^{\mathbb{R}_+}. \tag{19}$$

Note that the curvature term ξ behaves as a nonlinear damping term, slowing the system (from the perspective of the configuration space) as it approaches obstacles and vanishing when moving away from obstacles. Consequently, it biases the system toward curving along isocontours of the distance field. See Fig. 3 for a demonstration of these terms in isolation and in coordination with an obstacle repulsion potential.

D.4 Attractors

We detail a couple of attractor options here, including two metrics that we have used in practice. We sometimes find the more complex of these metrics works better on collision avoidance systems, since it expresses a desire for precision near the target while allowing orthogonal compliance further away giving some freedom for obstacle avoidance.

Notation Let $\mathbf{x}$ be the coordinate of the task space (e.g. the coordinate of the task space). We denote the inertia matrix as $\mathbf{M}(\mathbf{x})$, the forcing potential function as $\Phi(\mathbf{x})$, and the damping matrix as $\mathbf{B}(\mathbf{x}, \dot{\mathbf{x}})$. In this section, for designing attractors, we will focus on the special case where $\mathbf{M}(\mathbf{x}) = \mathbf{G}(\mathbf{x})$.

Acceleration-based attractors and GDSs In many cases, it is straightforward to design a task space behavior in isolation in terms of desired accelerations (either by hand or through planning) [16]. But for stability guarantees we want these systems to be GDSs. Specifically, suppose we have a motion policy given as $\mathbf{f}(\mathbf{x}, \dot{\mathbf{x}})$. Define $\ddot{\mathbf{x}}^d = \mathbf{f}(\mathbf{x}, \dot{\mathbf{x}})$ as a shorthand (i.e. the desired acceleration). We want to show that it can be written as

$$\ddot{\mathbf{x}}^d = -\mathbf{M}(\mathbf{x})^{-1}\Big(\nabla_{\mathbf{x}}\Phi(\mathbf{x}) + \boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) + \mathbf{B}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}}\Big) \tag{20}$$

for some $\mathbf{M}$, $\mathbf{B}$, and Φ, where $\boldsymbol{\xi}_{\mathbf{M}}$ is the associated curvature term of $\mathbf{M}$. We can view the above decomposition as three parts:

1. The desired acceleration generated by the potential: $-\mathbf{M}(\mathbf{x})^{-1}\nabla\Phi(\mathbf{x})$
2. The damping acceleration for stability: $-\mathbf{M}(\mathbf{x})^{-1}\mathbf{B}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}}$
3. The curvature acceleration for consistent behaviors: $-\mathbf{M}(\mathbf{x})^{-1}\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}})$

To bridge the connection between the GDS formulation in (20) and common motion policies given directly by $\mathbf{f}$, we consider particularly motion policy candidates that can be written in terms of

$$\ddot{\mathbf{x}}^d = -\nabla_{\mathbf{x}}\widetilde{\Phi}(\mathbf{x}) - \widetilde{\mathbf{B}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} - \mathbf{M}(\mathbf{x})^{-1}\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) \tag{21}$$

where $\widetilde{\Phi}$ is another potential function and $\widetilde{\mathbf{B}}(\mathbf{x}, \dot{\mathbf{x}})$ is another damping matrix. We show that it is possible to design $\widetilde{\Phi}$ and $\widetilde{\mathbf{B}}$ directly, and then choose some proper inertia matrix $\mathbf{M}(\mathbf{x})$ such that (21) can be written as the GDS (20) for some Φ and $\mathbf{B}$. That is, we show it possible to choose $\mathbf{M}$ such that

$$\nabla_x\Phi(\mathbf{x}) = \mathbf{M}(\mathbf{x})\nabla_{\mathbf{x}}\widetilde{\Phi}(\mathbf{x}) \qquad \text{and} \qquad \mathbf{B}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{M}(\mathbf{x})\widetilde{\mathbf{B}}(\mathbf{x}, \dot{\mathbf{x}})$$

and Φ is a potential function and $\mathbf{B}$ is positive definite (without the need to derive them in closed form). Moreover, we show that this strategy allows us to model some common acceleration-based attractors.

Motion policy candidates As a motivating example of (21), we consider the attractor proposed in [16]. Let h_V^α define a *soft-normalization* function

$$\theta_\alpha(\mathbf{v}) = \mathbf{v}/h_V^\alpha(\|\mathbf{v}\|). \tag{22}$$

with $h_V^\alpha(\gamma) = \frac{1}{\alpha}\log(e^{\alpha\gamma} + e^{-\alpha\gamma}) = \gamma + \frac{1}{\alpha}\log(1 + e^{-2\alpha\gamma})$ for some $\alpha > 0$, so that $\theta_\alpha(\mathbf{v})$ approaches $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ for larger $\mathbf{v}$, but approaches zero smoothly as $\mathbf{v} \to 0$. Without loss of generality, let us consider the center of attraction is at $\mathbf{x} = 0$. The attractor considered in [16] is given as

$$\ddot{\mathbf{x}}^d = \mathbf{f}_{\mathrm{a}}(\mathbf{x}, \dot{\mathbf{x}}) \coloneqq -\gamma_p \theta_\alpha(\mathbf{x}) - \gamma_d \dot{\mathbf{x}}, \tag{23}$$

for some $\gamma_p, \gamma_d > 0$.

Inspecting (23), we can see that it resembles the form (21) modulus the last curvature term $-\mathbf{M}(\mathbf{x})^{-1}\boldsymbol{\xi}(\mathbf{x}, \dot{\mathbf{x}})$. Indeed we can identify $\hat{\mathbf{B}} = \gamma_d \mathbf{I}$ and we show below the first term $\gamma_p \theta_\alpha(\mathbf{x}_0 - \mathbf{x})$ is a derivative of some potential function. We do so by showing its Jacobian is symmetric. Using the notation above, we have

$$\begin{aligned}
\frac{d}{d\mathbf{x}}\theta_\alpha(\mathbf{x}) &= \frac{d}{d\mathbf{x}}\left(h_V^\alpha(\|\mathbf{x}\|)^{-1}\mathbf{x}\right) \\
&= h_V^\alpha(\|\mathbf{x}\|)^{-1}\mathbf{I} + \mathbf{x}\left(-h_V^\alpha(\|\mathbf{x}\|)^{-2}\frac{dh_V^\alpha(s)}{ds}|_{s=\|\mathbf{x}\|}\frac{\partial}{\partial \mathbf{x}}\|\mathbf{x}\|\right) \\
&= h_V^\alpha(\|\mathbf{x}\|)^{-1}\mathbf{I} - \left(\|\mathbf{x}\|h_V^\alpha(\|\mathbf{x}\|)^{-2}\frac{dh_V^\alpha(s)}{ds}|_{s=\|\mathbf{x}\|}\right)\hat{\mathbf{x}}\hat{\mathbf{x}}^T.
\end{aligned}$$

Both terms are symmetric, so the Jacobian is symmeric and this vector field is the gradient of some potential function (say $\widetilde{\Phi}_{\mathrm{a}}^1$), although we do not attempt to derive the potential function in closed form here.

In some cases, it is potentially more convenient to start designing (21) with a known potential function such as

$$\begin{aligned}
\widetilde{\Phi}_{\mathrm{a}}^2(\mathbf{x}) &= \frac{1}{\eta}\log\left(e^{\eta\|\mathbf{x}\|} + e^{-\eta\|\mathbf{x}\|}\right) \\
&= \|\mathbf{x}\| + \frac{1}{\eta}\log\left(1 + e^{-2\eta\|\mathbf{x}\|}\right)
\end{aligned} \tag{24}$$

so the potential energy can be measured, where $\eta > 0$. This is a η-scaled softmax (η defines the effective smoothing radius at the origin) over $\|\mathbf{x}\|$ and $-\|\mathbf{x}\|$, and the second expression is a numerically robust version since $\|\mathbf{x}\| \geq 0$. Its negative

gradient is

$$\begin{aligned}\nabla_{\mathbf{x}}\widetilde{\Phi}^2(\mathbf{x}) &= \frac{1}{\alpha}\frac{1}{e^{\alpha\|\mathbf{x}\|}+e^{-\alpha\|\mathbf{x}\|}}\left(\alpha e^{\alpha\|\mathbf{x}\|}\hat{\mathbf{x}}-\alpha e^{-\alpha\|\mathbf{x}\|}\hat{\mathbf{x}}\right)\\ &= \left(\frac{e^{\alpha\|\mathbf{x}\|}-e^{-\alpha\|\mathbf{x}\|}}{e^{\alpha\|\mathbf{x}\|}+e^{-\alpha\|\mathbf{x}\|}}\right)\hat{\mathbf{x}}\\ &= \left(\frac{1-e^{-2\alpha\|\mathbf{x}\|}}{1+e^{-2\alpha\|\mathbf{x}\|}}\right)\hat{\mathbf{x}} = s_\alpha\big(\|\mathbf{x}\|\big)\hat{\mathbf{x}}, \end{aligned} \tag{25}$$

where $s_\alpha(0)=0$ and $s_\alpha(r)\to 1$ as $r\to\infty$. (25) again gives a numerically robust form since $\|\mathbf{x}\|\geq 0$. Below, we denote abstractly the potential as just $\widetilde{\Phi}$ so we're agnostic to the choice of $\widetilde{\Phi}^i_{\mathrm{a}}$, $i\in\{1,2\}$.

Metric options Suppose we have chosen some potential $\widetilde{\Phi}$ and some damping $\widetilde{\mathbf{B}}$. We next consider admissible metric/inertia matrices such that (21) can be written as (20). We first note that $\mathbf{M}=\mathbf{I}$ is an admissible choice (i.e. we recover $\Phi(\mathbf{x})=\widetilde{\Phi}$ and $\mathbf{B}=\widetilde{\mathbf{B}}$, provided $\widetilde{\mathbf{B}}$ is positive definite). But this choice is not ideal when we wish to combine multiple motion policies, because we recall that the design of $\mathbf{M}$ designates the importance of each motion policy. Therefore, we would not want to restrict ourselves to the trivial choice $\mathbf{M}=\mathbf{I}$.

Here we present a family of metric matrices that are non-trivial and meaningful in practice, and yet is *compatible* with the motion policy (21). Let us first define some useful functions to simplify the writing later on. Let $\alpha(\mathbf{x})=\exp(-\frac{\|\mathbf{x}\|^2}{2\sigma_\alpha^2})$ and $\gamma(\mathbf{x})=\exp(-\frac{\|\mathbf{x}\|^2}{2\sigma_\gamma^2})$ for some $\sigma_\alpha,\sigma_\gamma\in\mathbb{R}$. We define a weight function $w(\mathbf{x})=\gamma(\mathbf{x})w_u+(1-\gamma(\mathbf{x}))w_l$, for $0\leq w_l\leq w_u<\infty$. Equivalently, it can be written as $w(\mathbf{x})=\overset{\triangle}{w}\gamma(\mathbf{x})+w_l$ with $\overset{\triangle}{w}\coloneqq w_u-w_l$. Below we will need $\nabla_{\mathbf{x}}\log w(\mathbf{x})$, so we derive it here. Noting $w=\overset{\triangle}{w}\gamma+w_l$, we get

$$\nabla_{\mathbf{x}}\log w(\mathbf{x})=\frac{\nabla_{\mathbf{x}}(\overset{\triangle}{w}\gamma+w_l)}{w(\mathbf{x})}=\frac{\overset{\triangle}{w}}{w}\exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma_w^2}\right)\left(-\frac{1}{\sigma_w^2}\mathbf{x}\right)=-\frac{\gamma\overset{\triangle}{w}}{\sigma_w^2 w}\mathbf{x}.$$

We define two alternative metrics. The first metric trades off stretching the space in the direction toward the target when the robot is away from the goal, and becoming increasingly Euclidean when the robot is close to the goal:

$$\mathbf{M}_{\text{stretch}}=w(\mathbf{x})\Big(\big(1-\alpha(\mathbf{x})\big)\nabla_{\mathbf{x}}\widetilde{\Phi}\nabla_{\mathbf{x}}\widetilde{\Phi}^\top+(\alpha(\mathbf{x})+\epsilon)\mathbf{I}\Big), \tag{26}$$

where $\epsilon>0$ induces a baseline Euclidean metric used far from the target to fill out the metric's eigen-spectrum, and $\widetilde{\Phi}$ is the potential in (21). The second metric matrix is simply

$$\mathbf{M}_{\text{uni}}=w(\mathbf{x})\mathbf{I}. \tag{27}$$

We refer to these both generically as $\mathbf{M}$ below. Note again that we use $\mathbf{M}$ here rather than $\mathbf{G}$ since these metrics are velocity independent so that the inertia matrix $\mathbf{M}$ and the metric (typically denoted as $\mathbf{G}$) are the same.

Compatibility between metrics and potentials We show the two metrics $\mathbf{M}_{\text{stretch}}$ and $\mathbf{M}_{\text{uni}}$ above are compatible with $\widetilde{\Phi}^i_{\text{a}}$, $i \in \{1, 2\}$. For simplicity, let us denote them just as $\mathbf{M}$ and $\widetilde{\Phi}$. We will show that there exists a potential Φ such that $\nabla_{\mathbf{x}}\Phi = \mathbf{M}\nabla_{\mathbf{x}}\widetilde{\Phi}$. In fact, our result applies to potentials more general than $\widetilde{\Phi}^i_{\text{a}}$ for $i \in \{1, 2\}$. It applies to all potentials $\widetilde{\Phi}(\mathbf{x})$ such that $\nabla_{\mathbf{x}}\widetilde{\Phi}(\mathbf{x}) = \kappa(\|\mathbf{x}\|)\hat{\mathbf{x}}$ for some function $\kappa : \mathbb{R} \to \mathbb{R}$, which includes $\widetilde{\Phi}^i_{\text{a}}$ for $i \in \{1, 2\}$ as special cases.

We prove the existence by analyzing the Jacobian of $\mathbf{M}\nabla_{\mathbf{x}}\widetilde{\Phi}$ and showing that it is symmetric. We first note that a result of radial symmetry.

Lemma 6. *Let $\nabla_{\mathbf{x}}\widetilde{\Phi}(\mathbf{x}) = \kappa(\|\mathbf{x}\|)\hat{\mathbf{x}}$ for some $\kappa : \mathbb{R} \to \mathbb{R}$ operating on the distance to the origin, and let f be a differentiable function. Then the Jacobian matrix*

$$\frac{\partial}{\partial \mathbf{x}}\big(f\nabla_{\mathbf{x}}\widetilde{\Phi}\big) = f(\mathbf{x})\nabla^2_{\mathbf{x}}\widetilde{\Phi} + \kappa(\mathbf{x})f'(\|\mathbf{x}\|)\hat{\mathbf{x}}\hat{\mathbf{x}}^\top \tag{28}$$

is symmetric.

Proof. We first note that $\nabla f(\|\mathbf{x}\|) = f'(\|\mathbf{x}\|)\hat{\mathbf{x}}$ for all differentiable f. Then the results follow directly as the derivation below

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}\big(f\nabla_{\mathbf{x}}\widetilde{\Phi}\big) &= f(\mathbf{x})\nabla^2_{\mathbf{x}}\widetilde{\Phi} + \nabla\widetilde{\Phi}\nabla f^\top \\ &= f(\mathbf{x})\nabla^2_{\mathbf{x}}\widetilde{\Phi} + \gamma(\mathbf{x})f'(\|\mathbf{x}\|)\hat{\mathbf{x}}\hat{\mathbf{x}}^\top \end{aligned} \tag{29}$$

because the Hessian $\nabla^2\widetilde{\Phi}$ is symmetric. ■

Given Lemma 6, showing symmetry of the Jacobian of $\mathbf{M}\nabla_{\mathbf{x}}\widetilde{\Phi}$ is straightforward, because both $\widetilde{\Phi}$ considered satisfy $\nabla\widetilde{\Phi} = \kappa(\|\mathbf{x}\|)\hat{\mathbf{x}}$ for some κ. First, we consider $\mathbf{M}_{\text{uni}}$. We can write $\mathbf{M}_{\text{uni}}\nabla\widetilde{\Phi} = w(\mathbf{x})\nabla\widetilde{\Phi}$ and $w(\mathbf{x}) = \widetilde{w}(\|\mathbf{x}\|)$ with $\widetilde{w}(s) = \overset{\Delta}{w}\exp(-\frac{s^2}{2\sigma_\gamma}) + w_l$, so its Jacobian is symmetric. Similarly,

$$\begin{aligned}\mathbf{M}_{\text{stretch}}\nabla\widetilde{\Phi} &= w(\mathbf{x})\Big(\big(1-\alpha(\mathbf{x})\big)\nabla_{\mathbf{x}}\widetilde{\Phi}\nabla_{\mathbf{x}}\widetilde{\Phi}^T + \alpha(\mathbf{x})\Big)\nabla_{\mathbf{x}}\widetilde{\Phi} \\ &= w(\mathbf{x})\Big(\big(1-\alpha(\mathbf{x})\big)\|\nabla_{\mathbf{x}}\widetilde{\Phi}\|^2\nabla_{\mathbf{x}}\widetilde{\Phi} + \alpha(\mathbf{x})\nabla_{\mathbf{x}}\widetilde{\Phi}\Big) \\ &= \widetilde{w}\big(\|\mathbf{x}\|\big)h\big(\|\mathbf{x}\|\big)\nabla_{\mathbf{x}}\widetilde{\Phi},\end{aligned}$$

where $h(s) = (1-\widetilde{\alpha}(s))\kappa(s)^2 + \widetilde{\alpha}(s)$ with $\widetilde{\alpha}(s) = \exp(-\frac{s^2}{2\sigma_\alpha})$. This expression fits in the form considered in Lemma 6 and therefore it has a symmetric Jacobian.

Compatibility between metrics and damping The condition for the damping part is relative straightforward. We simply need to choose $\widetilde{\mathbf{B}}$ such that

$$\mathbf{B}(\mathbf{x}, \dot{\mathbf{x}}) = \mathbf{M}(\mathbf{x})\widetilde{\mathbf{B}}(\mathbf{x}, \dot{\mathbf{x}}) \succ 0$$

A sufficient condition is to set $\widetilde{\mathbf{B}}$ to share the same eigen-system as $\mathbf{M}$.

Effects of the curvature term We have provided conditions for compatibility between metrics, potentials, and damping. We now consider the effects of the curvature acceleration $-\mathbf{M}(\mathbf{x})^{-1}\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}})$

$$\ddot{\mathbf{x}}^d = -\nabla_{\mathbf{x}}\widetilde{\Phi}(\mathbf{x}) - \widetilde{\mathbf{B}}(\mathbf{x}, \dot{\mathbf{x}})\dot{\mathbf{x}} - \mathbf{M}(\mathbf{x})^{-1}\boldsymbol{\xi}_{\mathbf{M}}(\mathbf{x}, \dot{\mathbf{x}}) \tag{21}$$

due to our non-trivial choice of metric matrix.

For $\mathbf{M}_{\text{uni}} = w(\mathbf{x})\mathbf{I}$, this becomes

$$\begin{aligned}
\boldsymbol{\xi}_{\mathbf{M}} &= \frac{dw}{dt}\dot{\mathbf{x}} - \frac{1}{2}\nabla_{\mathbf{x}} w \|\dot{\mathbf{x}}\|^2 \\
&= \left(\dot{\mathbf{x}}^\top \nabla_{\mathbf{x}} w\right)\dot{\mathbf{x}} - \frac{1}{2}\nabla_{\mathbf{x}} w \|\dot{\mathbf{x}}\|^2 \\
&= \left(\left(\dot{\mathbf{x}}\dot{\mathbf{x}}^\top\right) - \frac{1}{2}\|\dot{\mathbf{x}}\|^2\right)\nabla_{\mathbf{x}} w \\
&= -\frac{1}{2}\|\dot{\mathbf{x}}\|^2 \left(\mathbf{I} - 2\ \hat{\dot{\mathbf{x}}}\hat{\dot{\mathbf{x}}}^T\right)\nabla_{\mathbf{x}} w
\end{aligned}$$

where $\hat{\dot{\mathbf{x}}} = \frac{\dot{\mathbf{x}}}{\|\dot{\mathbf{x}}\|}$. That gives

$$\begin{aligned}
-\mathbf{M}_{\text{uni}}^{-1}\boldsymbol{\xi}_{\mathbf{M}} &= \frac{1}{2}\|\dot{\mathbf{x}}\|^2 \left(\mathbf{I} - 2\ \hat{\dot{\mathbf{x}}}\hat{\dot{\mathbf{x}}}^T\right)\frac{\nabla w}{w} \\
&= \frac{1}{2}\|\dot{\mathbf{x}}\|^2 H^r_{\dot{\mathbf{x}}}\left[\nabla \log w\right],
\end{aligned}$$

where $H^r_{\mathbf{v}}[\mathbf{y}] = \left(\mathbf{I} - 2\ \hat{\mathbf{v}}\hat{\mathbf{v}}^T\right)\mathbf{y}$ is the Householder reflection of $\mathbf{y}$ across the plane normal to $\mathbf{v}$. In this case, it acts to align the system toward the goal and provides a bit of drag.

The derivatives of $\mathbf{M}_{\text{stretch}}$ are similar but more complex. We recommend a combination of finite-differencing and automatic differentiation to systematize the calculations.

Revisiting the attractor in [16] Let us revisit our motivating example

$$\ddot{\mathbf{x}}^d = \mathbf{f}_{\text{a}}(\mathbf{x}, \dot{\mathbf{x}}) \coloneqq -\gamma_p \theta_\alpha(\mathbf{x}) - \gamma_d \dot{\mathbf{x}}, \tag{23}$$

From using the results above, we see that (23) fits in the form in (21) but missing the curvature term $-\mathbf{M}^{-1}\boldsymbol{\xi}_{\mathbf{M}}$. As we show in Section 5.1, the curvature term provides correction for consistent behaviors and stability, which suggests that the original motion policy in [16] could lose stability in general (e.g. when the velocity is high). Nevertheless, from the above analysis, we show that if we add the curvature correction back, i.e.,

$$\ddot{\mathbf{x}} = \mathbf{f}_{\text{a}}(\mathbf{x}, \dot{\mathbf{x}}) - \mathbf{M}^{-1}\boldsymbol{\xi}_{\mathbf{M}}$$

then the system is provably stable.

D.5 Joint Limits

We adopt a similar approach to handling joint limits as [16], but here we modify the velocity dependent components of the metric to match our theoretical requirements for stability and fully derive the curvature terms introduced by the nonlinearities and velocity dependence. We emphasize that, due to the invariance of RMPs to reparameterization that results from our complete handling of curvature terms, the behavior of the joint limit RMPs and the way in which they interact with the rest of the system are independent of the specific implementation. That said, we derive an analogous result here to the one presented in [16] to show that these joint limit RMPs effectively scale the columns of each task space's Jacobian matrix to smoothly regulate their degrees of freedom as a function of joint limit proximity. In implementation, these RMPs can be treated as any other RMP.

Integrating RMPs with joint limits We first define a class of joint limit metrics that can be used in joint limit RMPs. We show, given a joint limit RMP, the RMP algebra defined in Section 3.4 can be seen as producing the same Jacobian modification as described in [16]. We present the result more generally here as a lemma. Note that as diagonal entries of $\mathbf{A}$ approach infinity, entries of $\mathbf{D}$ approach zero and the corresponding column of $\widetilde{\mathbf{J}}$ vanishes.

Lemma 7 (Effect of diagonal RMPs). *Let $\{(\ddot{\mathbf{x}}_i^d, \mathbf{M}_i)^{\mathcal{T}_i}\}_{i=1}^n$ be a collection of RMPs defined on task spaces $\mathcal{T}_i$. Let $\widetilde{\ddot{\mathbf{x}}}_i^d = \ddot{\mathbf{x}}_i^d - \dot{\mathbf{J}}_i\dot{\mathbf{q}}$ and let*

$$[\mathbf{M}\ddot{\mathbf{q}}^d, \mathbf{M}]^{\mathcal{C}} = \Big[\sum_i \mathbf{J}_i^T\mathbf{M}_i\widetilde{\ddot{\mathbf{x}}}_i^d,\ \sum_i \mathbf{J}_i^T\mathbf{M}_i\mathbf{J}_i\Big]^{\mathcal{C}} \tag{30}$$

denote their normal form pullback and combination to space $\mathcal{C}$ through task maps $\psi_i : \mathcal{C} \to \mathcal{T}_i$ with Jacobians $\mathbf{J}_i = \frac{\partial \psi_i}{\partial \mathbf{q}}$. Let $[\mathbf{A}\ddot{\mathbf{q}}_l^d, \mathbf{A}]^{\mathcal{C}}$ denote an RMP with diagonal[12] *a velocity-dependent metric $\mathbf{A}(\mathbf{q}, \dot{\mathbf{q}}) = \lambda\mathbf{D}(\mathbf{q}, \dot{\mathbf{q}})^{-2}$, where $\lambda > 0$. Then $[\mathbf{M}\ddot{\mathbf{q}}^d, \mathbf{M}]^{\mathcal{C}} + [\mathbf{A}\ddot{\mathbf{q}}_l^d, \mathbf{A}]^{\mathcal{C}}$ has metric $\widetilde{\mathbf{M}} = \mathbf{D}^{-1}\left(\sum_i \widetilde{\mathbf{J}}^T\mathbf{M}_i\widetilde{\mathbf{J}} + \lambda\mathbf{I}\right)\mathbf{D}^{-1}$ and motion policy*

$$\ddot{\mathbf{q}}_c^d = \mathbf{D}\left(\sum_i \widetilde{\mathbf{J}}^T\mathbf{M}_i\widetilde{\mathbf{J}} + \lambda\mathbf{I}\right)^{\dagger}\left(\sum_i \widetilde{\mathbf{J}}^T\mathbf{M}_i\widetilde{\ddot{\mathbf{x}}}_i^d + \lambda\mathbf{D}^{-1}\ddot{\mathbf{q}}_l^d\right) \tag{31}$$

$$= \mathbf{D}\left[\arg\min_{\ddot{\mathbf{q}}}\left(\frac{1}{2}\sum_i \|\widetilde{\ddot{\mathbf{x}}}_i^d - \widetilde{\mathbf{J}}_i\ddot{\mathbf{q}}\|_{\mathbf{M}_i}^2 + \frac{\lambda}{2}\|\widetilde{\ddot{\mathbf{q}}}_l^d - \ddot{\mathbf{q}}\|^2\right)\right], \tag{32}$$

with $\widetilde{\mathbf{J}} = \mathbf{J}\mathbf{D}$ and $\widetilde{\ddot{\mathbf{q}}}_l^d = \mathbf{D}^{-1}\ddot{\mathbf{q}}_l^d$.

[12] We choose this form for the diagonal dependent metric (without loss of generality since it's positive definite), to be convenient notationally below.

Proof. Writing out the sum we get

$$
\begin{aligned}
&[\mathbf{M}\ddot{\mathbf{q}}^d,\mathbf{M}]^{\mathcal{C}} + [\mathbf{A}\ddot{\mathbf{q}}_l^d, \mathbf{A}]^{\mathcal{C}} \\
&\quad = \left[\sum_i \mathbf{J}_i^T \mathbf{M}_i \widetilde{\ddot{\mathbf{x}}}_i^d + \lambda \mathbf{D}^{-2}\ddot{\mathbf{q}}_l^d,\ \sum_i \mathbf{J}_i^T \mathbf{M}_i \mathbf{J}_i + \lambda \mathbf{D}^{-2}\right]^{\mathcal{C}}. \\
&\quad = \left[\mathbf{D}^{-1}\Big(\sum_i \widetilde{\mathbf{J}}_i^T \mathbf{M}_i \widetilde{\ddot{\mathbf{x}}}_i^d + \lambda \mathbf{D}^{-1}\ddot{\mathbf{q}}_l^d\Big),\ \mathbf{D}^{-1}\Big(\sum_i \widetilde{\mathbf{J}}_i^T \mathbf{M}_i \widetilde{\mathbf{J}}_i + \lambda \mathbf{I}\Big)\mathbf{D}^{-1}\right]^{\mathcal{C}}.
\end{aligned}
$$

This gives the expression for the metric, and the motion policy can be obtained by `resolve`:

$$
\begin{aligned}
\widetilde{\ddot{\mathbf{q}}}_d &= \mathbf{D}\left(\sum_i \widetilde{\mathbf{J}}_i^T \mathbf{M}_i \widetilde{\mathbf{J}}_i + \lambda \mathbf{I}\right)^{\dagger}\left(\sum_i \widetilde{\mathbf{J}}_i^T \mathbf{M}_i \widetilde{\ddot{\mathbf{x}}}_i^d + \lambda \mathbf{D}^{-1}\ddot{\mathbf{q}}_l^d\right) \\
&= \mathbf{D}\ \texttt{resolve}\left(\Big[\sum_i \widetilde{\mathbf{J}}_i^T \mathbf{M}_i \widetilde{\ddot{\mathbf{x}}}_i^d + \lambda \mathbf{D}^{-1}\ddot{\mathbf{q}}_l^d, \sum_i \widetilde{\mathbf{J}}_i^T \mathbf{M}_i \widetilde{\mathbf{J}}_i + \lambda \mathbf{I}\Big]^{\mathcal{C}}\right),
\end{aligned}
$$

which is equivalent to the least squares form in (32). ■

A class of velocity-dependent joint limit metrics Here we develop a velocity dependent metric to represent joint limits. We construct it for each joint independently, denoting the joint angle by $q \in [l_l, l_u]$. Let $a(q, \dot{q})$ denote a one-dimensional velocity-dependent metric on q. We want $a \to \infty$ as q is close to the joint limit and $\dot{q}$ heads toward the joint limit. Such a metric can be constructed using a form related to that given in [16], choosing $a = b^{-2}$ for

$$
b = s\Big(\alpha_u d + (1-\alpha_u)1\Big) + (1-s)\Big(\alpha_l d + (1-\alpha_l)1\Big). \tag{33}
$$

with $s = \frac{q-l_l}{l_u-l_l}$, $d = 4s(1-s) = 4\left(\frac{q-l_l}{l_u-l_l}\right)\left(\frac{l_u-q}{l_u-l_l}\right)$, and velocity gates $\alpha_u = 1 - e^{-\dot{q}_+^2/(2\sigma^2)}$ and $\alpha_l = 1 - e^{-\dot{q}_-^2/(2\sigma^2)}$ for $\sigma > 0$. (Choosing $a = b^{-2}$ makes intuitive sense with regard to Lemma 7.) Since $s = \frac{q-l_l}{l_u-l_l}$ and $1-s = \frac{l_u-q}{l_u-l_l}$, s indicates whether q is close to l_u ($s \to 1$ as $q \to l_u$) while $1-s$ indicates whether it is close to l_l. Likewise, α_u indicates whether $\dot{q}$ is moving in a positive direction and α_l indicates a negative direction. Therefore, this equation encodes a smoothed binary logic that can be read "if close to the upper limit and moving in the positive direction use d, and if close to the lower limit and moving negatively d; in all other cases use 1." Said another way, "if close to either limit and moving toward it, use d, otherwise use 1." Note that $\sup d = 1$ and $d \to 0$ as $q \to \{l_l, l_u\}$, so $a = b^{-2}$ has the desired property discussed above. All that remains to be shown is that this choice of $a(q, \dot{q})$ satisfies the condition $\dot{q}\frac{\partial a}{\partial \dot{q}} \geq 0$ of Theorem 2.

Lemma 8. *The velocity-dependent metric defined by $a = b^{-2}$ with b given by* (33) *satisfies the sufficient condition of Theorem 2 for stability, i.e. $\frac{\partial a}{\partial \dot{q}}\dot{q} \geq 0$ for all $\dot{q}$.*

Proof. We start by noting

$$\frac{\partial a}{\partial \dot{q}}\dot{q} = \frac{\partial}{\partial \dot{q}} b^{-2}\dot{q} = -2b^{-3}\frac{\partial b}{\partial \dot{q}}\dot{q}. \tag{34}$$

Since $b \geq 0$ we have $\frac{\partial b}{\partial \dot{q}}\dot{q} \leq 0$ implies $\frac{\partial a}{\partial \dot{q}}\dot{q} \geq 0$. We can rearrange b to be more transparent to derivatives with respect to $\dot{q}$:

$$\begin{aligned} b &= s\Big(\alpha_u(d-1)+1\Big) + (1-s)\Big(\alpha_l(d-1)+1\Big) && (35)\\ &= -s(1-d)\alpha_u - (1-s)(1-d)\alpha_l + c, && (36)\\ &= -\gamma_u\alpha_u - \gamma_l\alpha_l + c, && (37)\end{aligned}$$

where c is independent of $\dot{q}$ and where $\gamma_u, \gamma_l \geq 0$ and both independent of $\dot{q}$. Therefore, we have

$$\frac{\partial b}{\partial \dot{q}}\dot{q} = -\gamma_u\frac{\partial \alpha_u}{\partial \dot{q}} - \gamma_l\frac{\partial \alpha_l}{\partial \dot{q}}. \tag{38}$$

Since $\frac{\partial \alpha_u}{\partial \dot{q}} > 0$ for $\dot{q} > 0$ and 0 otherwise, while $\frac{\partial \alpha_l}{\partial \dot{q}} < 0$ for $\dot{q} < 0$ and 0 otherwise, Equation 38 implies $\frac{\partial b}{\partial \dot{q}}\dot{q} \leq 0$ for all $\dot{q}$ and hence $\frac{\partial a}{\partial \dot{q}}\dot{q} \geq 0$ for all $\dot{q}$. ■

We note that there are other choices for joint limit metrics, including those used for obstacle avoidance. In fact, one way to create joint limit controllers would be to treat each joint limit as an obstacle. We choose to use the above limit controller due to its intuitive interpretation as a velocity-dependent modification of a controller designed in a space u with the relationship $q = (l_u - l_l)\sigma(u) + l_l$ with $\sigma(u) = 1/(1+e^{-u})$.

Motion policies for joint-limit avoidance The differential equation $\ddot{\mathbf{q}}_l^d$ from the joint limit RMPs (see Lemma 7) encodes the curvature terms from the metric $\mathbf{A}$. Denoting those as $\mathbf{A}^{-1}\boldsymbol{\xi}_{\mathbf{A}}$ with $\boldsymbol{\xi}_{\mathbf{A}} = \dot{\mathbf{A}}\dot{\mathbf{q}} - \frac{1}{2}\nabla_{\mathbf{q}}\left(\dot{\mathbf{q}}^T\mathbf{A}\dot{\mathbf{q}}\right) = \operatorname{diag}\left(\frac{1}{2}\frac{d\mathbf{A}_{ii}}{dq_i}\dot{q}_i^2\right)_i$, we often choose this differential equation to be

$$\ddot{\mathbf{q}}_l^d = \left(\eta_p(\mathbf{q}_0 - \mathbf{q}) - \eta_d\dot{\mathbf{q}}\right) - \mathbf{A}^{-1}\boldsymbol{\xi}_{\mathbf{A}}, \tag{39}$$

for $\eta_p, \eta_d \geq 0$. As shown in Lemma 7, this differential equation can be viewed as a transformation $\widetilde{\ddot{\mathbf{q}}}_l^d = \mathbf{D}^{-1}\ddot{\mathbf{q}}_l^d = \mathbf{A}^{\frac{1}{2}}\ddot{\mathbf{q}}_l^d$ in the final joint limit corrected expression. Since $\mathbf{A}$ becomes large near joint limits, this transformation effectively scales up the ith dimension of $\ddot{\mathbf{q}}_l^d$ when q_i nears a joint limit and $\dot{q}_i$ is headed toward it.

E Details of the Reaching-through-clutter Experiments

We give some details on the reaching experiments here.

E.1 Experimental method

We set up a collection of clutter-filled environments with cylindrical obstacles of varying sizes in simulation as depicted in Fig. 5, and tested the performance of RMPflow and two potential field methods on a modeled ABB YuMi robot.

Compared methods:

1. **RMPflow:** We implement RMPflow using the RMPs in Section 3.6 and detailed in Appendix D. In particular, we place collision-avoidance controllers on distance spaces $s_{ij} = d_j(\mathbf{x}_i)$, where $j = 1, \ldots, m$ indexes the world obstacle o_j and $i = 1, \ldots, n$ indexes the n control point along the robot's body. Each collision-avoidance controller uses a weight function $w_o(\mathbf{x})$ that ranges from 0 when the robot is far from the obstacle to $w_o^{\max} \gg 0$ when the robot is in contact with the obstacle's surface. Similarly, the attractor potential uses a weight function $w_a(\mathbf{x})$ that ranges from $w_a^{\min}$ far from the target to $w_a^{\max}$ close to the target.
2. **PF-basic:** This variant is a basic implementation of obstacle avoidance potential fields with dynamics shaping. We use the RMP framework to implement this variant by placing collision-avoidance controllers on the same body control points used in RMPflow but with isotropic metrics of the form $\mathbf{G}_o^{\text{basic}}(\mathbf{x}) = w_o^{\max}\mathbf{I}$ for each control point, with $w_o^{\max}$ matching the value RMPflow uses. Similarly, the attractor uses the same attractor potential as RMPflow, but with a constant isotropic metric with the form $\mathbf{G}_a^{\text{basic}}(\mathbf{x}) = w_a^{\max}\mathbf{I}$.
3. **PF-nonlinear:** This variant matches PF-basic in construction, except it uses a *nonlinear* isotropic metrics of the form $\mathbf{G}_o^{\text{nlin}}(\mathbf{x}_i) = w_o(\mathbf{x})\mathbf{I}$ and $\mathbf{G}_a^{\text{nlin}}(\mathbf{x}_i) = w_a(\mathbf{x})\mathbf{I}$ for obstacle-avoidance and attraction, respectively, using weight functions matching RMPflow.

A note on curvature terms: PF-basic uses constant metrics, so has no curvature terms; PF-nonlinear has nontrivial curvature terms arising from the spatially varying metrics, but we ignore them here to match common practice from the OSC literature.

Parameter scaling of PF-basic: Isotropic metrics do not express spacial directionality toward obstacles, and that leads to an inability of the system to effectively trade off the competing controller requirements. That conflict results in more collisions and increased instability. We, therefore, compare PF-basic under these baseline metric weights (matching RMPflow) with variants that incrementally strengthen collision avoidance controllers and C-space postural controllers ($f_{\mathcal{C}}(\mathbf{q}, \dot{\mathbf{q}}) = \gamma_p(\mathbf{q}_0 - \mathbf{q}) - \gamma_d\dot{\mathbf{q}}$) to improve these performance measures in the experiment. We use the following weight scalings (first entry denotes the obstacle metric scalar, and the second entry denotes the C-space metric scalar): "low" $(3, 10)$, "med" $(5, 50)$, and "high" $(10, 100)$.

Environments: We run each of these variants on 6 obstacle environments with 20 randomly sampled target locations each distributed on the opposite side of the obstacle field from the robot. Three of the environments use four smaller

obstacles (depicted in panel 3 of Fig. 5), and the remaining three environments used two large obstacles (depicted in panel 4 of Fig. 5). Each environment used the same 20 targets to avoid implicit sampling bias in target choice.

E.2 Performance measures

We report results in Fig. 4 in terms of mean and one standard deviation error bars calculated across the 120 trials for each of the following performance measures:[13]

1. *Time to goal ("time"):* Length of time, in seconds, it takes for the robot to reach a convergence state. This convergence state is either the target, or its best-effort local minimum. If the system never converges, as in the case of many potential field trials for infeasible problems, the trial times out after 5 seconds. This metric measures time-efficiency of the movement.
2. *C-space path length ("length"):* This is the total path length $\int \|\dot{\mathbf{q}}\| dt$ of the movement through the configuration space across the trial. This metric measures how economical the movement is. In many of the potential-field variants with lower weights, we see significant fighting among the controllers resulting in highly inefficient extraneous motions.
3. *Minimal achievable distance to goal ("goal distance"):* Measures how close, in meters, the system is able to get to the goal with its end-effector.
4. *Percent time in collision for colliding trials ("collision intensity"):* Given that a trial has a collision, this metric measures the fraction of time the system is in collision throughout the trial. This metric indicates the intensity of the collision. Low values indicate short grazing collisions while higher values indicate long term obstacle penetration.
5. *Fraction of trails with collisions ("collision failure"):* Reports the fraction of trials with any collision event. We consider these to be collision-avoidance controller failures.

E.3 Discussion

In Figure 4, we see that RMPflow outperforms each of these variants significantly, with some informative trends:

1. RMPflow never collides, so its collision intensity and collision failure values are 0.
2. The other techniques, progressing from no scaling of collision-avoidance and C-space controller weights to substantial scaling, show a profile of substantial collision in the beginning to fewer (but still non-zero) collision events in the end. But we note that improvement in collision-avoidance is achieved at the expense of time-efficiency and the robot's ability to reach the goal (it is too conservative).

[13] There is no guarantee of feasibility in planning problems in general, so in all cases, we measure performance relative to the performance of RMPflow, which is empirically stable and near optimal across these problems.

3. Lower weight scaling of both PF-basic and PF-nonlinear actually achieve some faster times and better goal distances, but that is because the system pushes directly through obstacles, effectively "cheating" during the trial. RMPflow remains highly economical with its best effort reaching behaviors while ensuring the trials remain collision-free.
4. Lower weight scalings of PF-basic are highly uneconomical in their motion reflective of their relative instability. As the C-space weights on the posture controllers increase, the stability and economy of motion increase, but, again, at the expense of time-efficiency and optimality of the final reach.
5. There is little empirical difference between PF-basic and PF-nonlinear indicating that the defining feature separating RMPflow from the potential field techniques is its use of a highly nonlinear metric that explicitly stretches the space in the direction of the obstacle as well as in the direction of the velocity toward the target. Those stretchings penalize deviations in the stretched directions during combination with other controllers while allowing variation along orthogonal directions. By being more explicit about how controllers should instantaneously trade off with one another, RMPflow is better able to mitigate the otherwise conflicting control signals.

Summary: Isotropic metrics do not effectively convey how each collision and attractor controller should trade off with one another, resulting in a conflict of signals that obscure the intent of each controller making simultaneous collision avoidance, attraction, and posture maintainence more difficult. Increasing the weights of the controllers can improve their effectiveness, but at the expence of descreased overall system performance. The resulting motions are slower and less effective in reaching the goal in spite of more stable behavior and fewer collisions. A key feature of RMPflow is its ability to leverage highly nonlinear metrics that better convey information about how controllers should trade off with one another, while retaining provable stability guarantees. In combination, these features result in efficient and economical obstacle avoidance behavior while reaching toward targets amid clutter.

F Details of integrated system

We demonstrate the integrated vision and motion system on two physical dual arm manipulation platforms: a Baxter robot from Rethink Robotics, and a YuMi robot from ABB. Footage of our fully integrated system (see start of Section 5 for the link) depicting tasks such as pick and place amid clutter, reactive manipulation of a cabinet drawers and doors with human interaction, *active* leadthrough with collision controllers running, and pick and place into a cabinet drawer.[14]

[14] We have also run the RMP portion of the system on an ABB IRB120 and a dual arm Kuka manipulation platform with lightweight collaborative arms. Only the two platforms mentioned here, the YuMi and the Baxter, which use the full motion and vision integration, are shown in the video for economy of space.

This full integrated system, shown in the supplementary video, uses the RMPs described in Section 3.6 (detailed in Appendix D) with a slight modification that the curvature terms are ignored. Instead, we maintain theoretical stability by using sufficient damping terms as described in Section 5.1 and by operating at slower speeds. Generalization of these RMPs between embodiments was anecdotally pretty consistent, although, as we demonstrate in our experiments, we would expect more empirical deviation at higher speeds. For these manipulation tasks, this early version of the system worked well as demonstrated in the video.

For visual perception, we leveraged consumer depth cameras along with two levels of perceptual feedback:

1. *Ambient world:* For the Baxter system we create a voxelized representation of the unmodeled ambient world, and use distance fields to focus the collision controllers on just the closest obstacle points surrounding the arms. This methodology is similar in nature to [13], except we found empirically that attending to only the closest point to a skeleton representation resulted in oscillation in concaved regions where distance functions might result in nonsmooth kinks. We mitigate this issue by finding the closest points to a *volume* around each control point, effectively smoothing over points of nondifferentiability in the distance field.
2. *Tracked objects:* We use the Dense Articulated Real-time Tracking (DART) system of [34] to track articulated objects in real time through manipulations. This system is able to track both the robot and environmental objects, such as an articulated cabinet, simulataneously to give accurate measurements of their relative configuration effectively obviating the need for explicit camera-world calibration. As long as the system is initialized in the general region of the object locations (where for the cabinet and the robot, that would mean even up to half a foot of error in translation and a similar scale of error in rotation), the DART optimizer will snap to the right configuration when turned on. DART sends information about object locations to the motion generation, and receives back information about expected joint configurations (priors) from the motion system generating a robust world representation usable in a number of practical real-world manipulation problems.

Each of our behaviors are decomposed as state machines that use visual feedback to detect transitions, including transitions to reaction states as needed to implement behavioral robustness. Each arm is represented as a separate robot for efficiency, receiving real-time information about other arm's current state enabling coordination. Both arms are programmed simultaneously using a high level language that provides the programmer a unified view of the surrounding world and command of both arms.