Geodesic Regression on the Grassmannian

Supplementary Material

Yi Hong[1], Roland Kwitt[2], Nikhil Singh[1], Brad Davis[3],
Nuno Vasconcelos[4] and Marc Niethammer[1,5]

[1] Department of Computer Science, UNC Chapel Hill, NC, United States
[2] Department of Computer Science, Univ. of Salzburg, Austria
[3] Kitware Inc., Carrboro, NC, United States
[4] Statistical and Visual Computing Lab, UCSD, CA, United States
[5] Biomedical Research Imaging Center, UNC Chapel Hill, NC, United States

This supplementary material contains technical details on the structure of the Grassmann manifold (**Section A**), our shooting strategy for Grassmannian geodesic regression (GGR, **Section B**), and the continuous piecewise GGR (**Section C**). The references to sections that appear in the paper [2] are marked as **[Paper, §xxx]**. The source code and further updates are also provided here: `https://yi_hong@bitbucket.org/yi_hong/ggr.git`.

A. Geodesic Equation / Variational Problem

To obtain the geodesic equation for the Grassmannian manifold $\mathcal{G}(p,n)$, we first consider curves $\mathbf{Y}(r) \in \mathcal{V}(p,n)$ and then refine the characterization of the tangents to tangents on $\mathcal{G}(p,n)$. The energy we want to minimize is

$$\begin{aligned} &E(\mathbf{Y}(r)) = \int_0^1 \operatorname{tr} \dot{\mathbf{Y}}(r)^\top \dot{\mathbf{Y}}(r) \, dr, \\ &\text{s.t. } \mathbf{Y}(0) = \mathbf{Y}_0, \; \mathbf{Y}(1) = \mathbf{Y}_1 \text{ and } \mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_p \; . \end{aligned} \tag{1}$$

In the paper, we use r_0 and r_1 as the integration bounds with $r_0 = 0$ and $r_1 = 1$. From [1], we know that for $\mathbf{Y}$ being a representer for an element $\mathcal{Y}$ of $\mathcal{G}(p,n)$, tangent vectors take the form

$$\Delta = \mathbf{Y}_\perp \mathbf{B} = (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C}, \tag{2}$$

where $\mathbf{C} \in \mathbb{R}^{n \times p}$ arbitrary. After dropping the argument r and the arguments of the energy in (1) for readability, the minimization problem becomes minimizing

$$E = \int_0^1 \operatorname{tr} \mathbf{C}^\top (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} \, dr = \int_0^1 \operatorname{tr} \mathbf{C}^\top (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} \, dr \tag{3}$$

such that $\dot{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C}$, and $\mathbf{Y}(0) = \mathbf{Y}_0$, $\mathbf{Y}(1) = \mathbf{Y}_1$, and $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_p$. Since we know that by enforcing the evolution equation for $\mathbf{Y}$, $\mathbf{Y}^\top\mathbf{Y} = \mathbf{I}_p$ will be fulfilled if it is fulfilled initially, it is sufficient to enforce $\mathbf{Y}(0)^\top\mathbf{Y}(0) = \mathbf{I}_p$, instead of enforcing it at every time point. To see this, note that

$$\frac{d}{dr}(\mathbf{Y}^\top\mathbf{Y}) = \dot{\mathbf{Y}}^\top\mathbf{Y} + \mathbf{Y}^\top\dot{\mathbf{Y}} = \mathbf{C}^\top(\mathbf{I} - \mathbf{Y}\mathbf{Y}^\top)\mathbf{Y} + \mathbf{Y}^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} = \mathbf{0}, \tag{4}$$

i.e., $\mathbf{Y}^\top\mathbf{Y}$ is constant under the evolution equation for $\mathbf{Y}$. Adding the constraints through Lagrangian multipliers, our objective is to find the critical points of the saddle point problem

$$\begin{aligned} E(\mathbf{C}, \mathbf{Y}, \lambda) = \int_0^1 \text{tr } \mathbf{C}^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} + \text{tr } \lambda^\top(\dot{\mathbf{Y}} - (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C})\ dr + \\ \text{tr } \lambda_c^\top(\mathbf{Y}(0)^\top\mathbf{Y}(0) - \mathbf{I}_p)\ . \end{aligned} \tag{5}$$

The variation is

$$\begin{aligned} \delta E = \int_0^1 \text{tr } (\delta\mathbf{C}^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} + \mathbf{C}^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\delta\mathbf{C}\ + \mathbf{C}^\top(-\delta\mathbf{Y}\mathbf{Y}^\top - \mathbf{Y}\delta\mathbf{Y}^\top)\mathbf{C}) + \\ \text{tr } (\delta\lambda^\top(\dot{\mathbf{Y}} - (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C}))\ + \text{tr } (\lambda^\top(\delta\dot{\mathbf{Y}} + \delta\mathbf{Y}\mathbf{Y}^\top\mathbf{C} + \mathbf{Y}\delta\mathbf{Y}^\top\mathbf{C} - (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\delta\mathbf{C}))\ dr + \\ \text{tr } \delta\lambda_c^\top(\mathbf{Y}(0)^\top\mathbf{Y}(0) - \mathbf{I}_p)\ + \text{tr } \lambda_c^\top(\delta\mathbf{Y}(0)^\top\mathbf{Y}(0) + \mathbf{Y}(0)^\top\delta\mathbf{Y}(0))\ . \end{aligned} \tag{6}$$

Collecting terms[1] yields

$$\begin{aligned} \delta E = \int_0^1 \text{tr } (2\mathbf{C}^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top) - \lambda^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top))\delta\mathbf{C}\ + \text{tr } \delta\lambda^\top(\dot{\mathbf{Y}} - (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C}) + \\ \text{tr } (-2\mathbf{Y}^\top\mathbf{C}\mathbf{C}^\top - \dot{\lambda}^\top + \mathbf{Y}^\top\mathbf{C}\lambda^\top + \mathbf{Y}^\top\lambda\mathbf{C}^\top)\delta\mathbf{Y}\ dr + \\ \text{tr } [\lambda^\top\delta\mathbf{Y}]_0^1 + \text{tr } \delta\lambda_c^\top(\mathbf{Y}(0)^\top\mathbf{Y}(0) - \mathbf{I}_p)\ + \text{tr } (\lambda_c^\top + \lambda_c)\mathbf{Y}(0)^\top\delta\mathbf{Y}(0)\ . \end{aligned} \tag{7}$$

Hence, we obtain the following optimality conditions

$$(2\mathbf{C}^\top - \lambda^\top)(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top) = \mathbf{0}, \tag{8}$$

$$\dot{\mathbf{Y}} - (\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} = \mathbf{0}, \tag{9}$$

$$-\dot{\lambda}^\top - 2\mathbf{Y}^\top\mathbf{C}\mathbf{C}^\top + \mathbf{Y}^\top\mathbf{C}\lambda^\top + \mathbf{Y}^\top\lambda\mathbf{C}^\top = \mathbf{0}, \tag{10}$$

subject to the boundary conditions $\mathbf{Y}(0) = \mathbf{Y}_0$, $\mathbf{Y}(1) = \mathbf{Y}_1$. **Note:** To obtain the geodesic equation, it is *not* necessary to obtain an expression for the dynamics of λ and $\mathbf{Y}$. Instead, we can simply multiply the evolution equation for $\mathbf{Y}$, i.e., (9) from the left with $\mathbf{Y}^\top$[2]. This yields

$$\mathbf{Y}^\top\dot{\mathbf{Y}} - \mathbf{Y}^\top(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)\mathbf{C} = \mathbf{Y}^\top\dot{\mathbf{Y}} = \mathbf{0}, \tag{11}$$

which upon differentiation becomes

$$\mathbf{Y}^\top\ddot{\mathbf{Y}} + \dot{\mathbf{Y}}^\top\dot{\mathbf{Y}} = \mathbf{0}\ . \tag{12}$$

This expression is different from the one given in [1]:

$$\ddot{\mathbf{Y}} + \mathbf{Y}\dot{\mathbf{Y}}^\top\dot{\mathbf{Y}} = \mathbf{0}\ . \tag{13}$$

[1] Useful relations are: integration by parts and tr $\mathbf{A} =$ tr $\mathbf{A}^\top$, tr $\mathbf{AB} =$ tr $\mathbf{BA}$, tr $\mathbf{ABC} =$ tr $\mathbf{CAB} =$ tr $\mathbf{BCA}$.
[2] This would not even have required setting up the variational problem in the first place.

However, substituting (13) into (12) yields

$$\mathbf{Y}^\top(-\mathbf{Y}\dot{\mathbf{Y}}^\top\dot{\mathbf{Y}}) + \dot{\mathbf{Y}}^\top\dot{\mathbf{Y}} = \mathbf{0} \ . \tag{14}$$

Hence, these equations are equivalent. Note also that by using the evolution equation (9) for $\mathbf{Y}$, and the relation between $\mathbf{C}$ and λ, given in (8), we obtain upon left multiplication of the evolution equation (10) for λ with $(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)$

$$(\mathbf{I}_n - \mathbf{Y}\mathbf{Y}^\top)(-\dot{\lambda} + \frac{1}{2}\lambda\lambda^\top\mathbf{Y}) = \mathbf{0} \ . \tag{15}$$

This determines the evolution of λ up to a projection onto the orthogonal complement of $\mathbf{Y}$. For a *free* initial condition $\mathbf{Y}(0)$, we obtain the additional optimality condition

$$-\lambda(0) + \mathbf{Y}(0)(\lambda_c + \lambda_c^\top) = \mathbf{0} \ . \tag{16}$$

By multiplying, from the left, with $\mathbf{Y}(0)^\top$ we obtain

$$\lambda_c + \lambda_c^\top = \mathbf{Y}(0)^\top\lambda(0) \ . \tag{17}$$

Backsubstitution (to eliminate λ_c) then yields the condition

$$-(\mathbf{I}_n - \mathbf{Y}(0)\mathbf{Y}(0)^\top)\lambda(0) = \mathbf{0} \ . \tag{18}$$

Hence, the gradient of E with respect to this initial condition $\mathbf{Y}(0)$ is given by

$$\begin{aligned} \nabla_{\mathbf{Y}(0)}E =& \nabla_{\mathbf{Y}(0)}d^2(\mathbf{Y}(0), \mathbf{Y}_1) \\ =& -(\mathbf{I}_n - \mathbf{Y}(0)\mathbf{Y}(0)^\top)\lambda(0) \\ \overset{(8)}{=}& -(\mathbf{I}_n - \mathbf{Y}(0)\mathbf{Y}(0)^\top)2\mathbf{C} \\ \overset{(9)}{=}& -2\dot{\mathbf{Y}}(0) \end{aligned} \tag{19}$$

which can be regarded as the gradient of the squared geodesic distance with respect to its current initial condition.

B. Derivation of Geodesic Shooting

In this section, we derive the shooting solution to the energy minimization problem **[Paper, §3.1]**. The shooting based energy for the geodesic determined by two representers, $\mathbf{Y}_0$ and $\mathbf{Y}_1$, is

$$E(\mathbf{Y}(r_0), \dot{\mathbf{Y}}(r_0)) = \ \alpha \ \mathrm{tr} \ \dot{\mathbf{Y}}(r_0)^\top\dot{\mathbf{Y}}(r_0) + \frac{1}{\sigma^2}d^2(\mathbf{Y}(r_1), \mathbf{Y}_1) \tag{20}$$

subject to the constraints

$$\begin{aligned} \mathbf{Y}(r_0)^\top\mathbf{Y}(r_0) &= \mathbf{I}_p, & (21)\\ \mathbf{Y}(r_0)^\top\dot{\mathbf{Y}}(r_0) &= \mathbf{0}, & (22)\\ \ddot{\mathbf{Y}}(r) + \mathbf{Y}(r)[\dot{\mathbf{Y}}(r)^\top\dot{\mathbf{Y}}(r)] &= \mathbf{0} \ . & (23) \end{aligned}$$

with balancing constants $\alpha \geq 0, \sigma > 0$. **Note:** As for a geodesic path the velocity has to be constant, we have replaced the integral, cf. (1), in the first term of (20) by its initial condition. To simplify computations, we replace the second order geodesic constraint by a system of first order. Specifically, introducing the auxiliary variables

$$\mathbf{X}_1(r) = \mathbf{Y}(r) \quad \text{and} \quad \mathbf{X}_2(r) = \dot{\mathbf{Y}}(r) \tag{24}$$

allows to write the shooting energy of (20) as

$$E(\mathbf{X}_1(r_0), \mathbf{X}_2(r_0)) = \alpha \text{ tr } \mathbf{X}_2(r_0)^\top \mathbf{X}_2(r_0) + \frac{1}{\sigma^2} d^2(\mathbf{X}_1(r_1), \mathbf{Y}_1) \tag{25}$$

such that

$$\begin{aligned}
\dot{\mathbf{X}}_1(r) &= \mathbf{X}_2(r), && (26)\\
\dot{\mathbf{X}}_2(r) &= -\mathbf{X}_1(r)[\mathbf{X}_2(r)^\top \mathbf{X}_2(r)], && (27)\\
\mathbf{X}_1(r_0)^\top \mathbf{X}_1(r_0) &= \mathbf{I}_p, && (28)\\
\mathbf{X}_1(r_0)^\top \mathbf{X}_2(r_0) &= \mathbf{0}\ . && (29)
\end{aligned}$$

Now, adding the constraints (26) – (29) through Lagrangian multipliers, we obtain the saddle-point problem associated with (20) as

$$\begin{aligned}
E(\mathbf{X}_1(r_0), \mathbf{X}_2(r_0)) = \alpha \text{ tr } \mathbf{X}_2(r_0)^\top \mathbf{X}_2(r_0) \ + \frac{1}{\sigma^2} d^2(\mathbf{X}_1(r_1), \mathbf{Y}_1) \ + \\
\int_{r_0}^{r_1} \text{tr } \lambda_1(r)^\top (\dot{\mathbf{X}}_1(r) - \mathbf{X}_2(r)) \ dr \ + \int_{r_0}^{r_1} \text{tr } \lambda_2(r)^\top (\dot{\mathbf{X}}_2(r) + \mathbf{X}_1(r)(\mathbf{X}_2(r)^\top \mathbf{X}_2(r))) \ dr \ + \\
\text{tr } \lambda_3^\top \mathbf{X}_1(r_0)^\top \mathbf{X}_2(r_0) + \text{tr } \lambda_4^\top (\mathbf{X}_1(r_0)^\top \mathbf{X}_1(r_0) - \mathbf{I}_p) \ .
\end{aligned} \tag{30}$$

Note: The Lagrangian multipliers $\lambda_1(r)$ and $\lambda_2(r)$ are dependent on the measurement variable r, whereas λ_3 and λ_4 are constants. Computing the associated variation yields

$$\begin{aligned}
\delta E(\mathbf{X}_1(r_0), \mathbf{X}_2(r_0)) = \alpha \text{ tr } 2\mathbf{X}_2(r_0)^\top \delta\mathbf{X}_2(r_0) \ + \frac{1}{\sigma^2} \nabla_{\mathbf{X}_1(r_1)} (d^2(\mathbf{X}_1(r_1), \mathbf{Y}_1))^\top \delta\mathbf{X}_1(r_1) \ + \\
\int_{r_0}^{r_1} \text{tr } \lambda_1(r)^\top (\delta\dot{\mathbf{X}}_1(r) - \delta\mathbf{X}_2(r)) \ + \text{tr } \delta\lambda_1(r)^\top (\dot{\mathbf{X}}_1(r) - \mathbf{X}_2(r)) \ dr \ + \\
\int_{r_0}^{r_1} \text{tr } \lambda_2(r)^\top (\delta\dot{\mathbf{X}}_2(r) + \delta\mathbf{X}_1(r)(\mathbf{X}_2(r)^\top \mathbf{X}_2(r)) \ + \mathbf{X}_1(r)(\delta\mathbf{X}_2(r)^\top \mathbf{X}_2(r) + \mathbf{X}_2(r)^\top \delta\mathbf{X}_2(r))) \ + \\
\text{tr } \delta\lambda_2(r)^\top (\dot{\mathbf{X}}_2(r) + \mathbf{X}_1(r)(\mathbf{X}_2(r)^\top \mathbf{X}_2(r))) \ dr \ + \\
\text{tr } \lambda_3^\top (\delta\mathbf{X}_1(r_0)^\top \mathbf{X}_2(r_0) + \mathbf{X}_1(r_0)^\top \delta\mathbf{X}_2(r_0)) \ + \text{tr } \delta\lambda_3^\top (\mathbf{X}_1(r_0)^\top \mathbf{X}_2(r_0)) \ + \\
\text{tr } \lambda_4^\top (\delta\mathbf{X}_1(r_0)^\top \mathbf{X}_1(r_0) + \mathbf{X}_1(r_0)^\top \delta\mathbf{X}_1(r_0)) \ + \text{tr } \delta\lambda_4^\top (\mathbf{X}_1(r_0)^\top \mathbf{X}_1(r_0) - \mathbf{I}_p) \ .
\end{aligned} \tag{31}$$

After collecting terms and integration by parts, we obtain

$$\begin{aligned}\delta E(\mathbf{X}_1(r_0),\mathbf{X}_2(r_0)) = \operatorname{tr}\,(-\lambda_1(r_0)^\top + \lambda_3\mathbf{X}_2(r_0)^\top + \lambda_4\mathbf{X}_1(r_0)^\top + \lambda_4^\top\mathbf{X}_1(r_0)^\top)\delta\mathbf{X}_1(r_0) +\\ \operatorname{tr}\,(\frac{1}{\sigma_2}\nabla_{\mathbf{X}_1(r_1)}(d^2(\mathbf{X}_1(r_1),\mathbf{Y}_1))^\top + \lambda_1(r_1)^\top)\delta\mathbf{X}_1(r_1) +\\ \operatorname{tr}\,(2\alpha\mathbf{X}_2(r_0)^\top - \lambda_2(r_0)^\top + \lambda_3^\top\mathbf{X}_1(r_0)^\top)\delta\mathbf{X}_2(r_0) + \operatorname{tr}\,\lambda_2(r_1)^\top\delta\mathbf{X}_2(r_1) +\\ \int_{r_0}^{r_1}\operatorname{tr}\,(-\dot{\lambda}_1(r)^\top + \mathbf{X}_2(r)^\top\mathbf{X}_2(r)\lambda_2(r)^\top)\delta\mathbf{X}_1(r)\,dr + \int_{r_0}^{r_1}\operatorname{tr}\,\delta\lambda_1(r)^\top(\dot{\mathbf{X}}_1(r) - \mathbf{X}_2(r))\,dr +\\ \int_{r_0}^{r_1}\operatorname{tr}\,(-\lambda_1(r)^\top - \dot{\lambda}_2(r)^\top + \mathbf{X}_1(r)^\top\lambda_2(r)\mathbf{X}_2(r)^\top + \lambda_2(r)^\top\mathbf{X}_1(r)\mathbf{X}_2(r)^\top)\delta\mathbf{X}_2(r)\,dr +\\ \int_{r_0}^{r_1}\operatorname{tr}\,\delta\lambda_2(r)^\top(\dot{\mathbf{X}}_2(r) + \mathbf{X}_1(r)(\mathbf{X}_2(r)^\top\mathbf{X}_2(r))\,dr +\\ \operatorname{tr}\,\delta\lambda_3^\top(\mathbf{X}_1(r_0)^\top\mathbf{X}_2(r_0)) + \operatorname{tr}\,\delta\lambda_4^\top(\mathbf{X}_1(r_0)^\top\mathbf{X}_1(r_0) - \mathbf{I}_p)\ .\end{aligned} \tag{32}$$

Hence, the following optimality conditions result:

$$\begin{cases}\dot{\mathbf{X}}_1(r) = & \mathbf{X}_2(r),\\ \dot{\mathbf{X}}_2(r) = & -\mathbf{X}_1(r)(\mathbf{X}_2(r)^\top\mathbf{X}_2(r)),\end{cases} \tag{33}$$

$$\begin{cases}\dot{\lambda}_1(r) = & \lambda_2(r)\mathbf{X}_2(r)^\top\mathbf{X}_2(r),\\ \dot{\lambda}_2(r) = & -\lambda_1(r) + \mathbf{X}_2(r)[\lambda_2(r)^\top\mathbf{X}_1(r) + \mathbf{X}_1(r)^\top\lambda_2(r)],\end{cases} \tag{34}$$

with boundary conditions and constraints

$$\mathbf{X}_1(r_0) = \mathbf{Y}(r_0) \tag{35}$$

$$\mathbf{X}_2(r_0) = \dot{\mathbf{Y}}(r_0) \tag{36}$$

$$\lambda_1(r_1) = -\frac{1}{\sigma^2}\nabla_{\mathbf{X}_1(r_1)}d^2(\mathbf{X}_1(r_1),\mathbf{Y}_1) \tag{37}$$

$$\lambda_2(r_1) = \mathbf{0}, \tag{38}$$

$$\mathbf{X}_1(r_0)^\top\mathbf{X}_2(r_0) = \mathbf{0}, \tag{39}$$

$$\mathbf{X}_1(r_0)^\top\mathbf{X}_1(r_0) = \mathbf{I}_p\ . \tag{40}$$

To compute the geodesic from a starting point $\mathbf{Y}(r_0)$, we want to determine $\dot{\mathbf{Y}}(r_0)$ instead of imposing it, hence $\mathbf{X}_2(r_0)$ is free which yields the condition from (32)

$$2\alpha\mathbf{X}_2(r_0) - \lambda_2(r_0) + \mathbf{X}_1(r_0)\lambda_3 = \mathbf{0}. \tag{41}$$

Multiplying this equation from the left by $\mathbf{X}_1(r_0)^\top$ yields

$$\lambda_3 = \mathbf{X}_1(r_0)^\top\lambda_2(r_0)\ . \tag{42}$$

and condition (41) becomes

$$2\alpha\mathbf{X}_2(r_0) - (\mathbf{I}_n - \mathbf{X}_1(r_0)\mathbf{X}_1(r_0)^\top)\lambda_2(r_0) = \mathbf{0}\ . \tag{43}$$

As this cannot be solved in closed form, we use an adjoint optimization problem and simply interpret (43) as the gradient with respect to the sought–for initial condition $\mathbf{X}_2(r_0) = \dot{\mathbf{Y}}(r_0)$

$$\nabla_{\mathbf{X}_2(r_0)} E = \; 2\alpha \mathbf{X}_2(r_0) - (\mathbf{I}_n - \mathbf{X}_1(r_0)\mathbf{X}_1(r_0)^\top)\lambda_2(r_0) \ . \tag{44}$$

In anticipation of the full regression formulation we also need the gradient with respect to $\mathbf{X}_1(r_0)$. If $\mathbf{X}_1(r_0)$ is free we obtain the additional optimality condition from (32)

$$-\lambda_1(r_0) + \mathbf{X}_2(r_0)\lambda_3^\top + \mathbf{X}_1(r_0)(\lambda_4 + \lambda_4^\top) = \mathbf{0} \ . \tag{45}$$

Multiplying this equation from the left by $\mathbf{X}_1(r_0)^\top$ yields

$$\lambda_4 + \lambda_4^\top = \mathbf{X}_1(r_0)^\top \lambda_1(r_0) \ . \tag{46}$$

Backsubstitution of (46) into (45) and using (42) for λ_3 then yields

$$-(\mathbf{I}_n - \mathbf{X}_1(r_0)\mathbf{X}_1(r_0)^\top)\lambda_1(r_0) + \mathbf{X}_2(r_0)\lambda_2(r_0)^\top \mathbf{X}_1(r_0) = \mathbf{0} \tag{47}$$

and consequently

$$\nabla_{\mathbf{X}_1(r_0)} E = -(\mathbf{I}_n - \mathbf{X}_1(r_0)\mathbf{X}_1(r_0)^\top)\lambda_1(r_0) \; + \mathbf{X}_2(r_0)\lambda_2(r_0)^\top \mathbf{X}_1(r_0) \ . \tag{48}$$

The derivation of geodesic shooting for multiple points is similar to the above procedures for two points, except for the data matching term, i.e., the second term in (20), which then includes distance terms for multiple measurements. This extension will not change the optimality conditions for initial conditions ($\mathbf{X}_1(r)$ and $\mathbf{X}_2(r)$) and the Lagrangian multipliers ($\lambda_1(r)$ and $\lambda_2(r)$), which are dependent on the variable r. But it will bring jumps to $\lambda_1(r)$ at the location of each measurement, r_i, which is similar to its boundary condition (37).

By shooting the optimality conditions (33) for the initial conditions forward, and the optimality conditions (34) for Lagrangian multipliers backward, as well as updating the initial conditions with gradients ((48) and (44)), we can solve the minimization problem for multiple points as shown in **[Paper, Algorithm 1]**.

C. Continuous Piecewise GGR

Using the adjoint method, it is easy to extend the Grassmannian geodesic regression (GGR) to a continuous piecewise version. In the following derivation, we mainly discuss the extension to two segments with the optimized location of the breakpoint; this is straightforward to be generalized to multiple segments, shown in Algorithms 1 and 2.

C.1. Optimal solution for fixed breakpoints

Given a set of data points on the Grassmannian, represented by $\mathbf{Y}_i$, and the breakpoints of the segments, we aim to find a continuous piecewise geodesic to fit these data measurements. To simplify the problem, we take two segments, *e.g.*, $[0, t-]$ and $[t+, 1]$, with t as the breakpoint. So the energy to be minimized is

$$E(\mathbf{X}_1(0), \mathbf{X}_2(0), \mathbf{X}_3(t+)) = \alpha\, tr\, \mathbf{X}_2(0)^\top \mathbf{X}_2(0) + \alpha\, tr\, \mathbf{X}_3(t+)^\top \mathbf{X}_3(t+) + \frac{1}{\sigma^2}\sum_{i=0}^{N-1} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i), \tag{49}$$

such that

$$\begin{cases} \dot{\mathbf{X}}_1 = \mathbf{X}_2, \quad \dot{\mathbf{X}}_2 = -\mathbf{X}_1\mathbf{X}_2^\top\mathbf{X}_2, \quad \mathbf{X}_1(0)^\top\mathbf{X}_1(0) = \mathbf{I}, \quad \mathbf{X}_1(0)^\top\mathbf{X}_2(0) = \mathbf{0}, \qquad [0, t-] \\ \dot{\mathbf{X}}_1 = \mathbf{X}_3, \quad \dot{\mathbf{X}}_3 = -\mathbf{X}_1\mathbf{X}_3^\top\mathbf{X}_3, \quad \mathbf{X}_1(t+)^\top\mathbf{X}_3(t+) = \mathbf{0}, \qquad [t+, 1] \\ \mathbf{X}_1(t-) = \mathbf{X}_1(t+) \ . \end{cases} \tag{50}$$

Here, $(\mathbf{X}_1(0), \mathbf{X}_2(0))$ is the initial point and the initial velocity for the first segment, and $(\mathbf{X}_1(t+), \mathbf{X}_3(t+))$ is the initial point and velocity for the second segment. By using the Lagrangian multipliers from λ_1 to λ_8, we can add all the eight constraints in (50) into the energy function for minimization. This yields

$$\begin{aligned} E &= \alpha\, tr\, \mathbf{X}_2(0)^\top\mathbf{X}_2(0) + \alpha\, tr\, \mathbf{X}_3(t+)^\top\mathbf{X}_3(t+) + \frac{1}{\sigma^2}\sum_{i=0}^{N-1} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i) \\ &+ \int_0^{t-} tr\, \lambda_1^\top(\dot{\mathbf{X}}_1 - \mathbf{X}_2)dr + \int_0^{t-} tr\, \lambda_2^\top(\dot{\mathbf{X}}_2 + \mathbf{X}_1\mathbf{X}_2^\top\mathbf{X}_2)dr + \int_{t+}^{1} tr\, \lambda_3^\top(\dot{\mathbf{X}}_1 - \mathbf{X}_3)dr \\ &+ \int_{t+}^{1} tr\, \lambda_4^\top(\dot{\mathbf{X}}_3 + \mathbf{X}_1\mathbf{X}_3^\top\mathbf{X}_3)dr + tr\, \lambda_5^\top(\mathbf{X}_1(0)^\top\mathbf{X}_1(0) - \mathbf{I}) \\ &+ tr\, \lambda_6^\top\mathbf{X}_1(0)^\top\mathbf{X}_2(0) + tr\, \lambda_7^\top\mathbf{X}_1(t+)^\top\mathbf{X}_3(t+) + tr\, \lambda_8^\top(\mathbf{X}_1(t-) - \mathbf{X}_1(t+)) \ . \end{aligned} \tag{51}$$

To obtain the optimality conditions for the optimization problem, we compute its variation, which results in

$$\begin{aligned} \delta E &= \alpha\, tr\, 2\mathbf{X}_2(0)^\top\delta\mathbf{X}_2(0) + \alpha\, tr\, 2\mathbf{X}_3(t+)^\top\delta\mathbf{X}_3(t+) + \frac{1}{\sigma^2}\sum_{i=0}^{N-1} \nabla_{\mathbf{X}_1(r_i)} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i)^\top\delta\mathbf{X}_1(r_i) \\ &+ \int_0^{t-} tr(\dot{\mathbf{X}}_1 - \mathbf{X}_2)^\top\delta\lambda_1 dr + tr\, \lambda_1(t-)^\top\delta\mathbf{X}_1(t-) - tr\, \lambda_1(0)^\top\delta\mathbf{X}_1(0) - \int_0^{t-} tr\, \dot{\lambda}_1^\top\delta\mathbf{X}_1 dr \\ &- \int_0^{t-} tr\, \lambda_1^\top\delta\mathbf{X}_2 dr + \int_0^{t-} tr(\dot{\mathbf{X}}_2 + \mathbf{X}_1\mathbf{X}_2^\top\mathbf{X}_2)^\top\delta\lambda_2 dr + tr\, \lambda_2(t-)^\top\delta\mathbf{X}_2(t-) - tr\, \lambda_2(0)^\top\delta\mathbf{X}_2(0) \\ &- \int_0^{t-} tr\, \dot{\lambda}_2^\top\delta\mathbf{X}_2 dr + \int_0^{t-} tr\, \mathbf{X}_2^\top\mathbf{X}_2\lambda_2^\top\delta\mathbf{X}_1 dr + \int_0^{t-} tr(\mathbf{X}_1^\top\lambda_2 + \lambda_2^\top\mathbf{X}_1)\mathbf{X}_2^\top\delta\mathbf{X}_2 dr \\ &+ \int_{t+}^{1} tr(\dot{\mathbf{X}}_1 - \mathbf{X}_3)^T\delta\lambda_3 dr + tr\, \lambda_3(1)^\top\delta\mathbf{X}_1(1) - tr\, \lambda_3(t+)^\top\delta\mathbf{X}_1(t+) - \int_{t+}^{1} tr\, \dot{\lambda}_3^\top\delta\mathbf{X}_1 dr \\ &- \int_{t+}^{1} tr\, \lambda_3^\top\delta\mathbf{X}_3 dr + \int_{t+}^{1} tr(\dot{\mathbf{X}}_3 + \mathbf{X}_1\mathbf{X}_3^\top\mathbf{X}_3)^\top\delta\lambda_4 dr + tr\, \lambda_4(1)^\top\delta\mathbf{X}_3(1) - tr\, \lambda_4(t+)^\top\delta\mathbf{X}_3(t+) \\ &- \int_{t+}^{1} tr\, \dot{\lambda}_4^\top\delta\mathbf{X}_3 dr + \int_{t+}^{1} tr\, \mathbf{X}_3^\top\mathbf{X}_3\lambda_4^\top\delta\mathbf{X}_1 dr + \int_{t+}^{1} tr(\mathbf{X}_1^\top\lambda_4 + \lambda_4^\top\mathbf{X}_1)\mathbf{X}_3^\top\delta\mathbf{X}_3 dr \\ &+ tr(\mathbf{X}_1(0)^\top\mathbf{X}_1(0) - \mathbf{I})^\top\delta\lambda_5 + tr(\lambda_5 + \lambda_5^\top)\mathbf{X}_1(0)^\top\delta\mathbf{X}_1(0) \\ &+ tr(\mathbf{X}_1(0)^\top\mathbf{X}_2(0))^\top\delta\lambda_6 + tr\, \lambda_6\mathbf{X}_2(0)^\top\delta\mathbf{X}_1(0) + tr\, \lambda_6^\top\mathbf{X}_1(0)^\top\delta\mathbf{X}_2(0) + \\ &+ tr(\mathbf{X}_1(t+)^\top\mathbf{X}_3(t+))^\top\delta\lambda_7 + tr\, \lambda_7\mathbf{X}_3(t+)^\top\delta\mathbf{X}_1(t+) + tr\, \lambda_7^\top\mathbf{X}_1(t+)^\top\delta\mathbf{X}_3(t+) \\ &+ tr(\mathbf{X}_1(t-) - \mathbf{X}_1(t+))^\top\delta\lambda_8 + tr\, \lambda_8^\top\delta\mathbf{X}_1(t-) - tr\, \lambda_8^\top\delta\mathbf{X}_1(t+) \ . \end{aligned} \tag{52}$$

By collecting terms we can simplify the variation as follows:

$$
\begin{aligned}
\delta E = &\int_0^{t-} tr(\dot{\mathbf{X}}_1 - \mathbf{X}_2)^\top \delta\lambda_1 dr + \int_0^{t-} tr(\dot{\mathbf{X}}_2 + \mathbf{X}_1\mathbf{X}_2^\top\mathbf{X}_2)^\top \delta\lambda_2 dr \\
&+ \int_0^{t-} tr(-\dot{\lambda}_1^\top + \mathbf{X}_2^\top\mathbf{X}_2\lambda_2^\top)\delta\mathbf{X}_1 dr + \int_0^{t-} tr(-\dot{\lambda}_2^\top - \lambda_1^\top + (\mathbf{X}_1^\top\lambda_2 + \lambda_2^\top\mathbf{X}_1)\mathbf{X}_2^\top)\delta\mathbf{X}_2 dr \\
&+ \int_{t+}^1 tr(\dot{\mathbf{X}}_1 - \mathbf{X}_3)^\top \delta\lambda_3 dr + \int_{t+}^1 tr(\dot{\mathbf{X}}_3 + \mathbf{X}_1\mathbf{X}_3^\top\mathbf{X}_3)^\top \delta\lambda_4 dr \\
&+ \int_{t+}^1 tr(-\dot{\lambda_3}^\top + \mathbf{X}_3^\top\mathbf{X}_3\lambda_4^\top)\delta\mathbf{X}_1 dr + \int_{t+}^1 tr(-\dot{\lambda}_4^\top - \lambda_3^\top + (\mathbf{X}_1^\top\lambda_4 + \lambda_4^\top\mathbf{X}_1)\mathbf{X}_3^\top)\delta\mathbf{X}_3 dr \\
&+ \frac{1}{\sigma^2}\sum_{i=0}^{N-1} \nabla_{\mathbf{X}_1(r_i)} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i)^\top \delta\mathbf{X}_1(r_i) + tr(-\lambda_1(0)^\top + (\lambda_5 + \lambda_5^\top)\mathbf{X}_1(0)^\top + \lambda_6\mathbf{X}_2(0)^\top)\delta\mathbf{X}_1(0) \\
&+ tr(2\alpha\mathbf{X}_2(0)^\top - \lambda_2^\top(0) + \lambda_6^\top\mathbf{X}_1(0)^\top)\delta\mathbf{X}_2(0) + tr(-\lambda_3(t+)^\top + \lambda_7\mathbf{X}_3(t+)^\top - \lambda_8^\top)\delta\mathbf{X}_1(t+) \\
&+ tr(2\alpha\mathbf{X}_3(t+)^\top - \lambda_4(t+)^\top + \lambda_7^\top\mathbf{X}_1(t+)^\top)\delta\mathbf{X}_3(t+) + tr(\lambda_1(t-)^\top + \lambda_8^\top)\delta\mathbf{X}_1(t-) \\
&+ tr\,\lambda_2(t-)^\top\delta\mathbf{X}_2(t-) + tr\,\lambda_3(1)^\top\delta\mathbf{X}_1(1) + tr\,\lambda_4(1)^\top\delta\mathbf{X}_3(1) + tr(\mathbf{X}_1(0)^\top\mathbf{X}_1(0) - \mathbf{I})^\top\delta\lambda_5 \\
&+ tr(\mathbf{X}_1(0)^\top\mathbf{X}_2(0))^\top\delta\lambda_6 + tr(\mathbf{X}_1(t+)^\top\mathbf{X}_3(t+))^\top\delta\lambda_7 + tr(\mathbf{X}_1(t-) - \mathbf{X}_1(t+))^\top\delta\lambda_8.
\end{aligned} \tag{53}
$$

Because the variations should vanish, we can get the forward system

$$
\begin{cases}
\dot{\mathbf{X}}_1 = \mathbf{X}_2, \quad \dot{\mathbf{X}}_2 = -\mathbf{X}_1\mathbf{X}_2^\top\mathbf{X}_2, & [0, t-] \\
\dot{\mathbf{X}}_1 = \mathbf{X}_3, \quad \dot{\mathbf{X}}_3 = -\mathbf{X}_1\mathbf{X}_3^\top\mathbf{X}_3, & [t+, 1]
\end{cases} \tag{54}
$$

with boundary conditions

$$
\begin{cases}
\mathbf{X}_1(0) = \mathbf{Y}(0), \quad \mathbf{X}_2(0) = \dot{\mathbf{Y}}(0), \\
\mathbf{X}_1(t+) = \mathbf{X}_1(t-), \quad \mathbf{X}_3(t+) = \dot{\mathbf{Y}}(t+)
\end{cases}. \tag{55}
$$

As we can see, for each segment, the state equations stay the same. The main difference lies in the values of the initial conditions, that is, the point, $\mathbf{X}_1$, remains continuous at the breakpoint in the forward integration, but the velocity, denoted by $\mathbf{X}_2$ and $\mathbf{X}_3$, will change according to the value at the breakpoint. $\mathbf{Y}(0)$, $\dot{\mathbf{Y}}(0)$, and $\dot{\mathbf{Y}}(t+)$ are unknown, and they will be updated using their corresponding gradients, which will be discussed later.

From (53), we can also obtain equations for the backward system

$$
\begin{cases}
\dot{\lambda}_1 = \lambda_2\mathbf{X}_2^\top\mathbf{X}_2, \quad \dot{\lambda}_2 = -\lambda_1 + \mathbf{X}_2(\lambda_2^\top\mathbf{X}_1 + \mathbf{X}_1^\top\lambda_2), & [0, t-] \\
\dot{\lambda}_3 = \lambda_4\mathbf{X}_3^\top\mathbf{X}_3, \quad \dot{\lambda}_4 = -\lambda_3 + \mathbf{X}_3(\lambda_4^\top\mathbf{X}_1 + \mathbf{X}_1^\top\lambda_4), & [t+, 1]
\end{cases} \tag{56}
$$

with boundary conditions

$$
\begin{cases}
\lambda_1(t-) = -\lambda_8, \quad \lambda_2(t-) = \mathbf{0}, \\
\lambda_3(1) = \mathbf{0}, \quad \lambda_4(1) = \mathbf{0}.
\end{cases} \tag{57}
$$

Here, λ_1 and λ_3 are equivalent since they are the adjoint variables for the point $\mathbf{X}_1$ at different segments. Similarly, λ_2 and λ_4 are the adjoint variables for the velocities, $\mathbf{X}_2$ and $\mathbf{X}_3$, at two segments. So the adjoint equations for each segment are also the same, but the boundary conditions are different at the breakpoint.

Since, $\mathbf{X}_1(0)$ is the initial point at time 0, and it should be free, we set

$$-\lambda_1(0) + \mathbf{X}_1(0)(\lambda_5 + \lambda_5^\top) + \mathbf{X}_2(0)\lambda_6^\top = \mathbf{0}. \tag{58}$$

Through left multiplication by $\mathbf{X}_1(0)^\top$, we can obtain

$$\lambda_5 + \lambda_5^\top = \mathbf{X}_1(0)^\top \lambda_1(0). \tag{59}$$

Similarly, $\mathbf{X}_2(0)$ is the initial velocity at time 0 and should be free, we set

$$2\alpha\mathbf{X}_2(0) - \lambda_2(0) + \mathbf{X}_1(0)\lambda_6 = \mathbf{0}. \tag{60}$$

Left multiplication by $\mathbf{X}_1(0)^\top$ gives

$$\lambda_6 = \mathbf{X}_1(0)^\top \lambda_2(0). \tag{61}$$

By substituting the values of λ_5 and/or λ_6 into the left side of (58) and (60), we can obtain the gradients for $\mathbf{X}_1(0)$ and $\mathbf{X}_2(0)$ as

$$\nabla_{\mathbf{X}_1(0)} E = -(\mathbf{I} - \mathbf{X}_1(0)\mathbf{X}_1(0)^\top)\lambda_1(0) + \mathbf{X}_2(0)\lambda_2(0)^\top \mathbf{X}_1(0), \tag{62}$$

$$\nabla_{\mathbf{X}_2(0)} E = 2\alpha\mathbf{X}_2(0) - (\mathbf{I} - \mathbf{X}_1(0)\mathbf{X}_1(0)^\top)\lambda_2(0). \tag{63}$$

Besides, $\mathbf{X}_3(t+)$ is the initial velocity at the breakpoint, and it also should be free. To obtain its gradient, we set

$$2\alpha\mathbf{X}_3(t+) - \lambda_4(t+) + \mathbf{X}_1(t+)\lambda_7 = \mathbf{0}. \tag{64}$$

Left multiplication by $\mathbf{X}_1(t+)^\top$ results in

$$\lambda_7 = \mathbf{X}_1(t+)^\top \lambda_4(t+). \tag{65}$$

So by substituting the value of λ_7 into the left side of (64), we can obtain the gradient for $\mathbf{X}_3(t+)$:

$$\nabla_{\mathbf{X}_3(t+)} E = 2\alpha\mathbf{X}_3(t+) - (\mathbf{I} - \mathbf{X}_1(t+)\mathbf{X}_1(t+)^\top)\lambda_4(t+). \tag{66}$$

The gradients, $\nabla_{\mathbf{X}_2(0)} E$ and $\nabla_{\mathbf{X}_3(t+)} E$, share the same formulation because $\mathbf{X}_2(0)$ and $\mathbf{X}_3(t+)$ are the initial velocity of each segment, $\lambda_2(0)$ and $\lambda_4(t+)$ are the corresponding adjoint variables, and $\mathbf{X}_1(0)$ and $\mathbf{X}_1(t+)$ are the initial point of each segment. That is, we can use one formulation to compute the gradient of the initial velocity for all segments.

There is still one thing left, computing the value of λ_8. Because $\mathbf{X}_1(t+)$ is the initial point at the breakpoint, but it has to be the same with the ending point of the previous piece, this means it should not be free, and no gradient needs to be computed. But we can use it to compute the value of λ_8 that is required for computing the boundary condition for λ_1:

$$-\lambda_3(t+) + \mathbf{X}_3(t+)\lambda_7^\top - \lambda_8 = \mathbf{0} \tag{67}$$

Algorithm 1: Continuous Piecewise Regression on the Grassmannian

Data: $\{(r_i, \mathbf{Y}_i)\}_{i=0}^{N-1}$, α, σ^2, and breakpoints $\{t_j\}_{j=1}^{K}$
Result: $\mathbf{Y}(r_0)$, $\dot{\mathbf{Y}}(r_0)$, and $\{\dot{\mathbf{Y}}(t_j)\}_{j=1}^{K}$
Set initial $\mathbf{Y}(r_0)$, $\dot{\mathbf{Y}}(r_0)$, and $\{\dot{\mathbf{Y}}(t_j)\}_{j=1}^{K}$.
Set $t_0 = r_0$. // the minimal value of $\{r_i\}_{i=0}^{N-1}$
Set $t_{K+1} = r_{N-1}$. // the maximal value of $\{r_i\}_{i=0}^{N-1}$
while *not converged* **do**
  for j := *0 to* K **do**
    Solve $\begin{cases} \dot{\mathbf{X}}_1 = \mathbf{X}_2, \; \mathbf{X}_1(t_j) = \mathbf{Y}(t_j), \\ \dot{\mathbf{X}}_2 = -\mathbf{X}_1(\mathbf{X}_2^\top \mathbf{X}_2), \; \mathbf{X}_2(t_j) = \dot{\mathbf{Y}}(t_j) \end{cases}$ forward for $r \in [t_j, t_{j+1}]$.
  end
  $\lambda_8 = \mathbf{0}$.
  for j := K *to 0* **do**
    Solve $\begin{cases} \dot{\lambda}_1 = \lambda_2 \mathbf{X}_2^\top \mathbf{X}_2, \; \lambda_1(t_{j+1}+) = -\lambda_8, \\ \dot{\lambda}_2 = -\lambda_1 + \mathbf{X}_2(\lambda_2^\top \mathbf{X}_1 + \mathbf{X}_1^\top \lambda_2), \; \lambda_2(t_{j+1}) = \mathbf{0} \end{cases}$ backward for $r \in [t_j, t_{j+1}]$
    with jump conditions

$$\lambda_1(r_i-) = \lambda_1(r_i+) - \frac{1}{\sigma^2} \nabla_{\mathbf{X}_1(r_i)} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i), \quad t_j \le r_i \le t_{j+1}$$

    and $\nabla_{\mathbf{X}_1(r_i)} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i)$ is computed as in Section A. For multiple measurements at a given r_i, the jump conditions for each measurement are added up.
    Compute boundary condition for λ_1 at the breakpoint t_j:

$$\lambda_8 = -\lambda_1(t_j+) + \mathbf{X}_2(t_j)\lambda_2(t_j)^\top \mathbf{X}_1(t_j).$$

    Compute gradient with respect to the initial velocity at the breakpoint t_j (or at the point r_0 when $j = 0$):

$$\nabla_{\dot{\mathbf{Y}}(t_j)} E = 2\alpha \mathbf{X}_2(t_j) - (\mathbf{I}_n - \mathbf{X}_1(t_j)\mathbf{X}_1(t_j)^\top)\lambda_2(t_j).$$

  end
  Compute gradient with respect to the initial point:

$$\nabla_{\mathbf{Y}(r_0)} E = -(\mathbf{I}_n - \mathbf{X}_1(r_0)\mathbf{X}_1(r_0)^\top)\lambda_1(r_0-) + \mathbf{X}_2(r_0)\lambda_2(r_0)^\top \mathbf{X}_1(r_0).$$

  Use a line search with these gradients to update $\mathbf{Y}(r_0)$, $\dot{\mathbf{Y}}(r_0)$, and $\{\dot{\mathbf{Y}}(t_j)\}_{j=1}^{K}$, as described in **[Paper, Algorithm 2]**.
end

By substituting (65) into (67), we can compute the λ_8 as

$$\lambda_8 = -\lambda_3(t+) + \mathbf{X}_3(t+)\lambda_4(t+)^\top \mathbf{X}_1(t+) \tag{68}$$

The last term $\frac{1}{\sigma^2}\sum_{i=0}^{N-1} \nabla_{\mathbf{X}_1(r_i)} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i)^\top \delta \mathbf{X}_1(r_i)$ will be used to derive the jump conditions shown in Algorithm 1.

Compared to the standard geodesic regression, the continuous piecewise regression shares the same forward and backward systems. The only difference is that in the piecewise regression, its boundary conditions for both forward and backward systems change at the breakpoints. Specifically, there are two modifications in the continuous piecewise regression:

1) In the forward system, the initial point at the breakpoint remains unchanged, the same with the ending point of the previous segment, but the initial velocity at the breakpoint needs to be updated based on the corresponding gradient, i.e., the gradient with respect to $\mathbf{X}_3(t+)$.

2) In the backward system, the boundary condition for the Lagrangian multiplier of the initial velocity remains 0 at the breakpoint as before, but the boundary condition for the Lagrangian multiplier of the initial point is no longer 0, but will be set to $-\lambda_8$, which is computed using (68).

Since for each segment all the equations remain the same compared to the standard geodesic regression with only slight changes at the boundary conditions, it is straightforward to extend the two segments to multiple segments, as shown in Algorithm 1.

C.2. Optimal breakpoints for two segments

In Section C.1, the continuous piecewise regression is performed with an assumption that the breakpoint is known. In this section, our goal is to optimize the location of the breakpoint for the continuous piecewise GGR. To address this problem, we rewrite the cost function (see (51)) using the lim operator, which yields

$$
\begin{aligned}
E = \alpha\, tr\, \mathbf{X}_2(0)^\top \mathbf{X}_2(0) + \alpha \lim_{t \to t+} tr\, \mathbf{X}_3(t)^\top \mathbf{X}_3(t) + \frac{1}{\sigma^2} \sum_{i=0}^{N-1} d^2(\mathbf{X}_1(r_i), \mathbf{Y}_i) \\
+ \lim_{t \to t-} \int_0^t tr\, \lambda_1^\top (\dot{\mathbf{X}}_1 - \mathbf{X}_2) dr + \lim_{t \to t-} \int_0^t tr\, \lambda_2^\top (\dot{\mathbf{X}}_2 + \mathbf{X}_1 \mathbf{X}_2^\top \mathbf{X}_2) dr + \lim_{t \to t+} \int_t^1 tr\, \lambda_3^\top (\dot{\mathbf{X}}_1 - \mathbf{X}_3) dr \\
+ \lim_{t \to t+} \int_t^1 tr\, \lambda_4^\top (\dot{\mathbf{X}}_3 + \mathbf{X}_1 \mathbf{X}_3^\top \mathbf{X}_3) dr + tr\, \lambda_5^\top (\mathbf{X}_1(0)^\top \mathbf{X}_1(0) - \mathbf{I}) + tr\, \lambda_6^\top \mathbf{X}_1(0)^\top \mathbf{X}_2(0) \\
+ \lim_{t \to t+} tr\, \lambda_7^\top \mathbf{X}_1(t)^\top \mathbf{X}_3(t) + tr\, \lambda_8^\top (\lim_{t \to t-} \mathbf{X}_1(t) - \lim_{t \to t+} \mathbf{X}_1(t))
\end{aligned}
\tag{69}
$$

To find the optimal position for the breakpoint, we first compute the gradient for updating. By taking the derivative of the energy with respect to t, and based on the fact that tr and lim are linear operators, we can obtain the gradient with respect to t as

$$
\begin{aligned}
\frac{dE}{dt} &= 0 + \alpha\, tr \lim_{t\to t+} \frac{d\mathbf{X}_3(t)^\top \mathbf{X}_3(t)}{dt} + 0 + \lim_{t\to t-} tr\, \lambda_1^\top (\dot{\mathbf{X}}_1 - \mathbf{X}_2) + \lim_{t\to t-} tr\, \lambda_2^\top (\dot{\mathbf{X}}_2 + \mathbf{X}_1 \mathbf{X}_2^\top \mathbf{X}_2) \\
&\quad - \lim_{t\to t+} tr\, \lambda_3^\top (\dot{\mathbf{X}}_1 - \mathbf{X}_3) - \lim_{t\to t+} tr\, \lambda_4^\top (\dot{\mathbf{X}}_3 + \mathbf{X}_1 \mathbf{X}_3^\top \mathbf{X}_3) + 0 + 0 \\
&\quad + tr \lim_{t\to t+} \frac{d\lambda_7^\top \mathbf{X}_1(t)^\top \mathbf{X}_3(t)}{dt} + tr\, \lambda_8^\top \lim_{t\to t-} \frac{d\mathbf{X}_1(t)}{dt} - tr\, \lambda_8^\top \lim_{t\to t+} \frac{d\mathbf{X}_1(t)}{dt} \\
&\overset{(50)}{=} \alpha\, tr \lim_{t\to t+} \frac{d\mathbf{X}_3(t)^\top \mathbf{X}_3(t)}{dt} + tr \lim_{t\to t+} \frac{d\lambda_7^\top \mathbf{X}_1(t)^\top \mathbf{X}_3(t)}{dt} \\
&\quad + tr\, \lambda_8^\top \lim_{t\to t-} \frac{d\mathbf{X}_1(t)}{dt} - tr\, \lambda_8^\top \lim_{t\to t+} \frac{d\mathbf{X}_1(t)}{dt} \\
&= \alpha\, tr \lim_{t\to t+} (\dot{\mathbf{X}}_3(t)^\top \mathbf{X}_3(t) + \mathbf{X}_3(t)^\top \dot{\mathbf{X}}_3(t)) + tr \lim_{t\to t+} (\lambda_7^\top \dot{\mathbf{X}}_1(t)^\top \mathbf{X}_3(t) + \lambda_7^\top \mathbf{X}_1(t)^\top \dot{\mathbf{X}}_3(t)) \\
&\quad + tr\, \lambda_8^\top \dot{\mathbf{X}}_1(t-) - tr\, \lambda_8^\top \dot{\mathbf{X}}_1(t+) \\
&\overset{(50)}{=} -\alpha\, tr\, \mathbf{X}_3(t+)^\top \mathbf{X}_3(t+)(\mathbf{X}_1(t+)^\top \mathbf{X}_3(t+)) - \alpha\, tr(\mathbf{X}_3(t+)^\top \mathbf{X}_1(t+))\mathbf{X}_3(t+)^\top \mathbf{X}_3(t+) \\
&\quad + tr\, \lambda_7^\top (\mathbf{I} - \mathbf{X}_1(t+)^\top \mathbf{X}_1(t+))\mathbf{X}_3(t+)^\top \mathbf{X}_3(t+) + tr\, \lambda_8^\top \mathbf{X}_2(t-) - tr\, \lambda_8^\top \mathbf{X}_3(t+) \\
&\overset{(50)}{=} tr\, \lambda_8^\top (\mathbf{X}_2(t-) - \mathbf{X}_3(t+)).
\end{aligned} \tag{70}
$$

This indicates that the breakpoint of two segments is optimized based on the jumps of the initial velocity at the breakpoint, which could be extended to multiple segments. In Algorithm 2, we use an iterative optimization strategy to find the optimal locations for multiple breakpoints.

Algorithm 2: Continuous Piecewise GGR with Optimal Breakpoints

---

Data: $\{(r_i, \mathbf{Y}_i)\}_{i=0}^{N-1}$, α, σ^2, and K (the number of breakpoints)
Result: $\mathbf{Y}(r_0)$, $\dot{\mathbf{Y}}(r_0)$, $\{t_j, \mathbf{Y}(t_j)\}_{j=1}^{K}$
Set initial $\{t_j\}_{j=1}^{K}$, evenly distributed within (r_0, r_{N-1}).
Set initial $\mathbf{Y}(r_0)$, $\dot{\mathbf{Y}}(r_0)$, and $\{\dot{\mathbf{Y}}(t_j)\}_{j=1}^{K}$.
while *not converged* **do**
    Update $\mathbf{Y}(r_0)$, $\dot{\mathbf{Y}}(r_0)$, and $\{\dot{\mathbf{Y}}(t_j)\}_{j=1}^{K}$ using Algorithm 1, and save the values of $\lambda_8(t_j)$, $\mathbf{X}_2(t_j-)$, and $\mathbf{X}_2(t_j+)$ at each breakpoint $\{t_j\}_{j=1}^{K}$.
    for *j := 1 to K* **do**
        Compute gradient with respect to t at the breakpoint t_j

$$
\left.\frac{dE}{dt}\right|_{t=t_j} = tr\, \lambda_8(t_j)^\top (\mathbf{X}_2(t_j-) - \mathbf{X}_2(t_j+)).
$$

    end
    Use a line search with the gradients to update $\{t_j\}_{j=1}^{K}$.
end

References

[1] A. Edelman, T. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998. 1, 2

[2] Y. Hong, R. Kwitt, N. Singh, B. Davis, N. Vasconcelos, and M. Niethammer. Geodesic regression on the Grassmannian. In *ECCV*, 2014. 1