NA 568 - Winter 2022

RKHS Registration

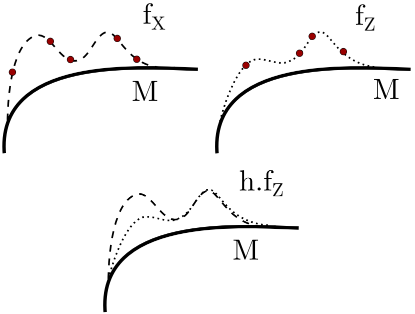
Maani Ghaffari

March 22, 2022

UNIVERSITY OF MICHIGAN

Nonparametric Continuous Sensor Registration

A new mathematical framework for sensor registration that enables nonparametric joint semantic and geometric representation of continuous functions using data.



- Clark, W., Ghaffari, M. and Bloch, A., 2021. Nonparametric Continuous Sensor Registration. Journal of Machine Learning Research, 22, pp.1-50. https://arxiv.org/pdf/2001.04286.pdf

This work will deal with point clouds on *Manifolds*. A manifold is something that locally looks like Euclidean space.

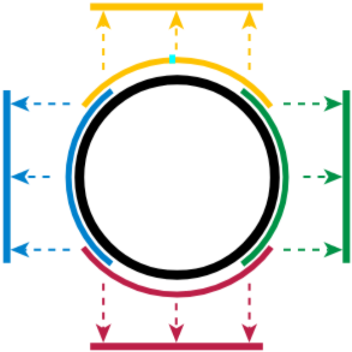


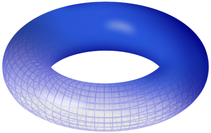
Figure: https://en.wikipedia.org/wiki/Manifold

Theoretical Preliminaries - Manifolds

Many common objects are manifolds.

1. Every Euclidean space, $\mathbb{R}^n$.
2. The 2-sphere, S^2

3. The Torus T^2



https://www.jpl.nasa.gov/edu/teach/activity/ocean-world-earth-globe-toss-game/
https://en.wikipedia.org/wiki/Torus

An inner product generalizes the notion of a dot product.

Definition

Given a vector space V, an inner product is a map $\langle \cdot,\cdot \rangle : V \times V \to \mathbb{R}$ such that

1. $\langle x,y \rangle = \langle y,x \rangle$ (symmetric)
2. $\langle \alpha x,y \rangle = \alpha \langle x,y \rangle$ and $\langle x + z,y \rangle = \langle x,y \rangle + \langle z,y \rangle$ (linear)
3. $\langle x,x \rangle > 0$ if $x \neq 0$ (positive-definite)

Then the pair, $(V,\langle \cdot,\cdot \rangle)$ is called an inner product space.

Recall that $\langle x,y \rangle = 0$ if they are orthogonal and the bigger $\langle x,y \rangle$ is, the more parallel they are.

Examples: Nonparametric Continuous Sensor Registration

$X = \{x_i\}$

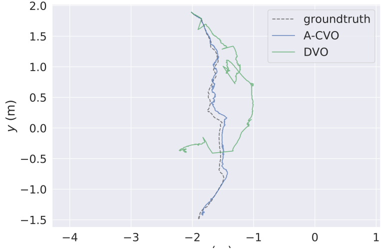
Target Cloud

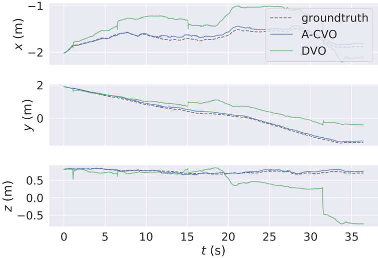
$Z = \{z_j\}$

Source Cloud

$hZ = \{hz_j\}$

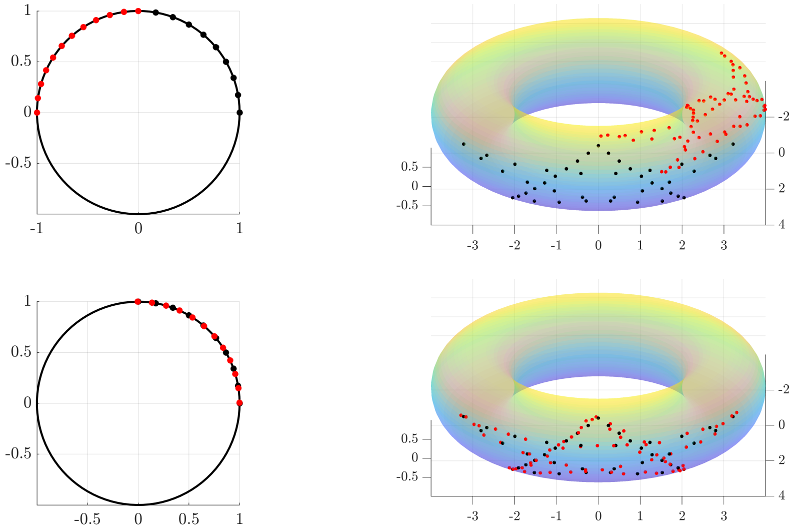
$h \in \mathrm{SE}_3$





- Maani Ghaffari, William Clark, Anthony Bloch, Ryan M. Eustice, and Jessy W. Grizzle. "Continuous Direct Sparse Visual Odometry from RGB-D Images," in Proceedings of Robotics: Science and Systems, Freiburg, Germany, June 2019.
- Tzu-Yuan Lin, William Clark, Ryan M. Eustice, Jessy W. Grizzle, Anthony Bloch, and Maani Ghaffari. "Adaptive Continuous Visual Odometry from RGB-D Images." arXiv preprint arXiv:1910.00713, 2019.

Examples: Nonparametric Continuous Sensor Registration



Examples: Nonparametric Continuous Sensor Registration

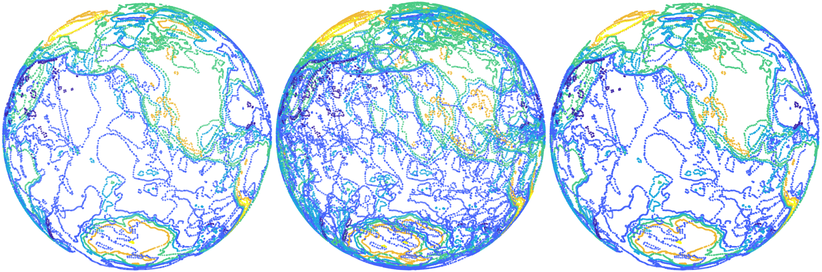


Figure: Left (Reference): The image of the contour map of the Earth. Center (Perturbed): The plot of X along with $A \cdot Z$. Right (Aligned): The plot of X along with $T \cdot A \cdot Z$.

Examples: Nonparametric Continuous Sensor Registration

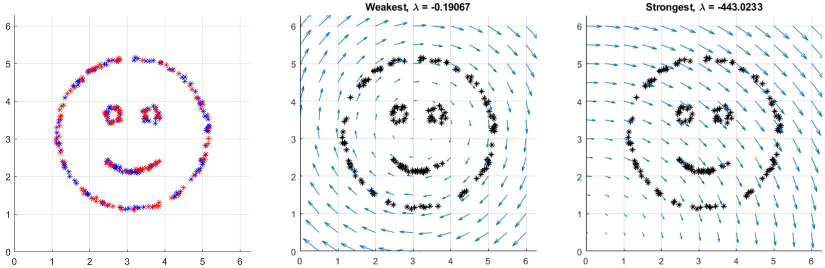


Figure: Left: The images of the two point clouds. The blue stars represent X while the red are Z. Center: The vector field corresponding to the eigenvector of the Hessian with the weakest eigenvalue. Right: The vector field corresponding to the eigenvector of the Hessian with the strongest eigenvalue.

Definition

A sensor registration problem is given by a 5-tuple: $(\mathcal{G}, M, \varphi, \langle\cdot,\cdot\rangle_{\mathfrak{g}}, k)$ where

(G1) $\mathcal{G}$ is a Lie group,

(G2) M is a smooth manifold,

(G3) $\varphi : \mathcal{G} \to \mathrm{Diff}(M)$ is a smooth group action,

(G4) $\langle\cdot,\cdot\rangle_{\mathfrak{g}}$ is an inner product on $\mathfrak{g} = \mathrm{Lie}(\mathcal{G})$, and

(G5) $k : M \times M \to \mathbb{R}$ is a symmetric positive definite function, called the kernel,

Definition

A sensor registration problem is given by a 5-tuple: $(\mathcal{G}, M, \varphi, \langle\cdot,\cdot\rangle_{\mathfrak{g}}, k)$ where

(G1) $\mathcal{G}$ is a Lie group,

(G2) M is a smooth manifold,

(G3) $\varphi : \mathcal{G} \to \mathrm{Diff}(M)$ is a smooth group action,

(G4) $\langle\cdot,\cdot\rangle_{\mathfrak{g}}$ is an inner product on $\mathfrak{g} = \mathrm{Lie}(\mathcal{G})$, and

(G5) $k : M \times M \to \mathbb{R}$ is a symmetric positive definite function, called the kernel,

while the information is given by a 3-tuple: $(\mathcal{I}, (X, \ell_X), (Z, \ell_Z))$ where

(I1) $\mathcal{I}$ is an inner product space, called the information space,

(I2) $X \subset M$; target cloud and $\ell_X : X \to \mathcal{I}$ is its label,

(I3) $Z \subset M$; source cloud and $\ell_Z : Z \to \mathcal{I}$ is its label.

- The information (G1)-(G5) is required to build the general form of the gradient.

- The information (I1)-(I3) encodes the actual clouds which is subsequently plugged into the gradient.

- Choose a symmetric positive definite function $k : M \times M \to \mathbb{R}$ to be the kernel of a Reproducing Kernel Hilbert Space (RKHS), $\mathcal{H}$.

- This allows us to turn the point clouds to functions via

$$f_X(\cdot) := \sum_{x_i \in X} \ell_X(x_i) k(\cdot, x_i), \quad f_Z(\cdot) := \sum_{z_j \in Z} \ell_Z(z_j) k(\cdot, z_j).$$

We can now define the inner product of f_X and f_Z by

$$\langle f_X, f_Z \rangle_{\mathcal{H}} := \sum_{\substack{x_i \in X \\ z_j \in Z}} \langle \ell_X(x_i), \ell_Z(z_j) \rangle_{\mathcal{I}} \cdot k(x_i, z_j).$$

Problem

The problem of aligning the point clouds can now be phrased as maximizing the scalar products of f_X and $h.f_Z$, i.e., we want to solve

$$\arg\max_{h \in \mathcal{G}} F(h), \quad F(h) := \langle f_X, h.f_Z \rangle_{\mathcal{H}}.$$

Problem

The problem of aligning the point clouds can now be phrased as maximizing the dot product of f_X and $h.f_Z$, i.e.

$$\max_{h \in \mathcal{G}} F(h), \quad F(h) := \langle f_X, h.f_Z \rangle_{\mathcal{H}}.$$

$$f_X(\cdot) := \sum_{x_i \in X} \ell_X(x_i) k(\cdot, x_i),$$

$$h.f_Z(\cdot) := \sum_{z_j \in Z} \ell_Z(z_j) k(\cdot, h^{-1} z_j),$$

$$\langle f_X, h.f_Z \rangle_{\mathcal{H}} := \sum_{\substack{x_i \in X \\ z_j \in Z}} \langle \ell_X(x_i), \ell_Z(z_j) \rangle_{\mathcal{I}} \cdot k(x_i, h^{-1} z_j).$$

For the "point cloud dot product"

$$\langle f_X, h.f_Z\rangle_{\mathcal{H}} := \sum_{\substack{x_i \in X \\ z_j \in Z}} \langle \ell_X(x_i), \ell_Z(z_j)\rangle_{\mathcal{I}} \cdot k(x_i, h^{-1} z_j).$$

we will take

$$k(x,y) = \sigma^2 \exp\left(\frac{-\|x-y\|^2}{2\ell^2}\right)$$

and the term $\langle \ell_X(x_i), \ell_Z(z_j)\rangle_{\mathcal{I}}$ is the dot product of the color of x_i and z_j.
For simplicity, we will take

$$c_{ij} := \langle \ell_X(x_i), \ell_Z(z_j)\rangle_{\mathcal{I}}$$

Recall that we need to maximize the function

$$F(h) = \sum_{\substack{x_i \in X \\ z_j \in Z}} c_{ij} \cdot k(x_i, h^{-1} z_j).$$

This will be done by gradient ascent,

$$\dot{h} = \nabla F(h).$$

The steady-state of this ODE is where $\nabla F(h) = 0$ which will be a (local) maximum.

Suppose we want to maximize some (differentiable) function $f : \mathbb{R}^n \to \mathbb{R}$.
If we differentiate it, ∇f is the vector that points in the direction of maximal growth.
Therefore, following this path will lead to a (local) maximum.

$$\dot{x} = \nabla f(x)$$

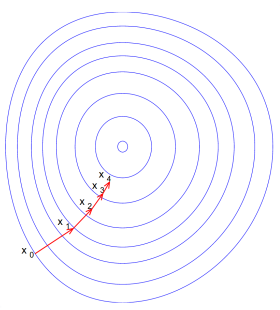


Figure: https://en.wikipedia.org/wiki/Gradient_descent

Given ∇F, we can follow the ode $\dot{h} = \nabla F(h)$ to find the maximum. Explicitly, we have the ode:

$$\dot{R} = R\omega,$$
$$\dot{T} = Rv.$$

$$\omega = \frac{1}{a^2\ell^2} \sum_{\substack{x_i \in X \\ z_j \in Z}} c_{ij} \cdot k(x_i, h^{-1}z_j) \cdot \left(x_i \times (h^{-1}z_j) \right),$$

$$v = \frac{1}{b^2\ell^2} \sum_{\substack{x_i \in X \\ z_j \in Z}} c_{ij} \cdot k(x_i, h^{-1}z_j) \cdot \left(x_i - h^{-1}z_j \right).$$

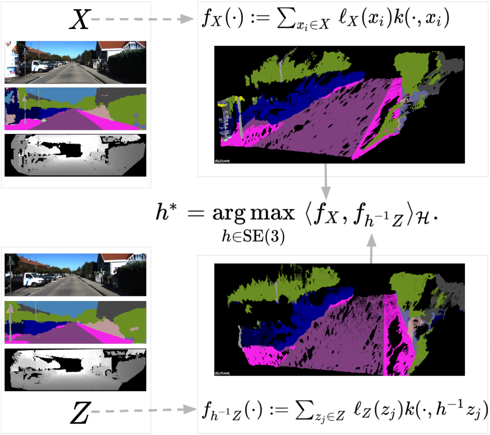
Let $\exp : \mathfrak{se}(3) \to \mathrm{SE}(3)$ be the Lie exponential map (this is just the matrix exponential). Then our integration step is merely

$$\begin{aligned}
\begin{bmatrix} R_{k+1} & T_{k+1} \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} R_k & T_k \\ 0 & 1 \end{bmatrix} \cdot \exp \begin{bmatrix} t\omega & tv \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} R_k & T_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta R & \Delta T \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} R_k \Delta R & R_k \Delta T + T_k \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

Therefore, we arrive at

$$\begin{aligned}
R_{k+1} &= R_k \Delta R \\
T_{k+1} &= R_k \Delta T + T_k
\end{aligned}$$

Semantic CVO



Zhang, R., et al., 2021, May. A new framework for registration of semantic point clouds from stereo and RGB-D cameras. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 12214-12221).
`https://arxiv.org/abs/2012.03683`
https://www.youtube.com/watch?v=M3XZOAWyu04

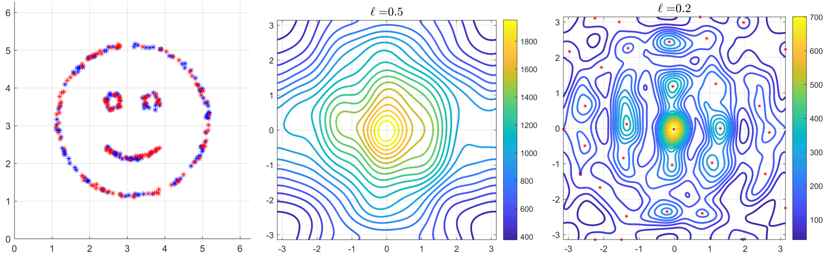


Figure: Left: The images of the two point clouds (now viewed in $[0,2\pi)^2 \sim T^2$ rather than $\mathbb{R}^2$). The blue stars represent X while the red are Z. Center: A contour plot of $F : T^2 \to \mathbb{R}$ where $\ell = 0.5$. Right: A contour plot of $F : T^2 \to \mathbb{R}$ where $\ell = 0.2$. A total of 33 local maxima are found in this picture. A video of the effect of varying ℓ is available at https://youtu.be/ETr6-c0VapQ.

Reduction to the Weighted Least Squares

- Without loss of generality, we will use the Gaussian (Squared Exponential) kernel based on this distance function, $d(\cdot,\cdot)$: $k(x,y) = \sigma^2 \exp\left(-d(x,y)^2\right)$.

- Further, let us define $w_{ij} := \sigma^2 \langle \ell_X(x_i), \ell_Z(z_j) \rangle_{\mathcal{I}}$ and $d_{ij} := d(x_i, z_j)$. Then we have

$$\langle f_X, f_Z \rangle_{\mathcal{H}} = \sum_{\substack{x_i \in X \\ z_j \in Z}} w_{ij} \cdot \exp\left(-d_{ij}^2\right).$$

Reduction to the Weighted Least Squares

Assumption

The assumptions used in the least squares problem are:

1. *The point clouds X and Z have the same number of points.*
2. *The point cloud Z is ordered such that the matching indices are corresponding measurements.*
3. *The residuals are only computed between corresponding measurements.*

By applying the above-mentioned assumptions, we get

$$\langle f_X, f_Z \rangle_{\mathcal{H}} \approx \sum_i w_i \cdot \exp\left(-d_i^2\right) .$$

Finally, the $\exp$ can be approximated using a Taylor expansion as

$$\exp\left(-d_i^2\right) \approx 1 - d_i^2 + \cdots .$$

Problem

The problem of aligning the point clouds can now be rephrased as the following weighted least squares form

$$\arg\min_{h \in \mathcal{G}} S(h), \quad S(h) := \sum_i w_i \cdot d(x_i, hz_i)^2.$$

Open-Source Software

- https://github.com/MaaniGhaffari/cvo-rgbd
- https://github.com/UMich-CURLY/unified_cvo
- https://github.com/MaaniGhaffari/c-sensor-registration
- https://github.com/minghanz/c3d

Readings

- Clark, W., Ghaffari, M. and Bloch, A., 2021. Nonparametric Continuous Sensor Registration. Journal of Machine Learning Research, 22, pp.1-50.
- Zhang, R., Lin, T.Y., Lin, C.E., Parkison, S.A., Clark, W., Grizzle, J.W., Eustice, R.M. and Ghaffari, M., 2021, May. A new framework for registration of semantic point clouds from stereo and RGB-D cameras. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 12214-12221).
- Zhu, M., Ghaffari, M. and Peng, H., 2022, January. Correspondence-free point cloud registration with SO (3)-equivariant implicit shape representations. In Conference on Robot Learning (pp. 1412-1422).
- Zhu, M., Ghaffari, M., Zhong, Y., Lu, P., Cao, Z., Eustice, R.M. and Peng, H., 2020. Monocular depth prediction through continuous 3D loss. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 10742-10749).