Approximating solutions to partial differential equations with physics-informed reproducing kernels

Valdemar Skou Knudsen

University of Copenhagen

Presentation outline

- Introduce RKHSs and basic properties.
- Discuss use in PDE solution, with main reference being Stepaniants 2023.
- Discuss what type of physics-informed constraints I consider, and methodology for incorporating them in kernel design.
- Present and discuss experiments and results, future outlook.

Reproducing Kernel Hilbert Spaces

Definition and characterization

Definition

Let $\mathcal{F}(X)$ denote the set of complex-valued functions on some non-empty background set X. A reproducing kernel Hilbert space (RKHS) is then a tuple $(\mathcal{H}, K)$ where $\mathcal{H} \subseteq \mathcal{F}(E)$ is a Hilbert space of functions, and $K : X \times X \to \mathbb{C}$ satisfies the reproducing property: for all $p \in X$ and $f \in \mathcal{H}$ we have that $K_p := K(\cdot, p) \in \mathcal{H}$ and $f(p) = \langle f, K_p \rangle_{\mathcal{H}}$.

The Moore-Aronszajn theorem gives an explicit characterization of the Hilbert space:

$$\mathcal{H} = \overline{\left\{ x \mapsto \sum_{i=1}^{n} \alpha_i K(x, x_i) \mid n \in \mathbb{N}, \forall i = 1, \ldots, n : \alpha_i \in, x_i \in X \right\}}$$

More on RKHSs

- A function $\hat{f} \in \mathcal{H}$ minimizing a RKHS penalized risk function on a finite training sample $((x_1, y_1), \ldots, (x_n, y_n))$, can be expressed as a finite kernel expansion:

$$\hat{f}(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i).$$

  Sometimes, an explicit formulation of weights α can be given (i.e. least squares). If a more complicated loss is employed, or if the unknown function is learned indirectly, iterative optimization schemes are required.
- A robust tool with wide applicability.
- Kernel transformations for specialized uses are viable.

Physics-Informed RKHS

Motivation

- Goal: use RKHSs to approximate functions u that are also solutions to PDEs, i.e. (linear PDE):

$$\sum_{|\alpha| \leq d} a_\alpha(x) D^\alpha u = f(x), \quad x \in \Omega, \tag{1}$$

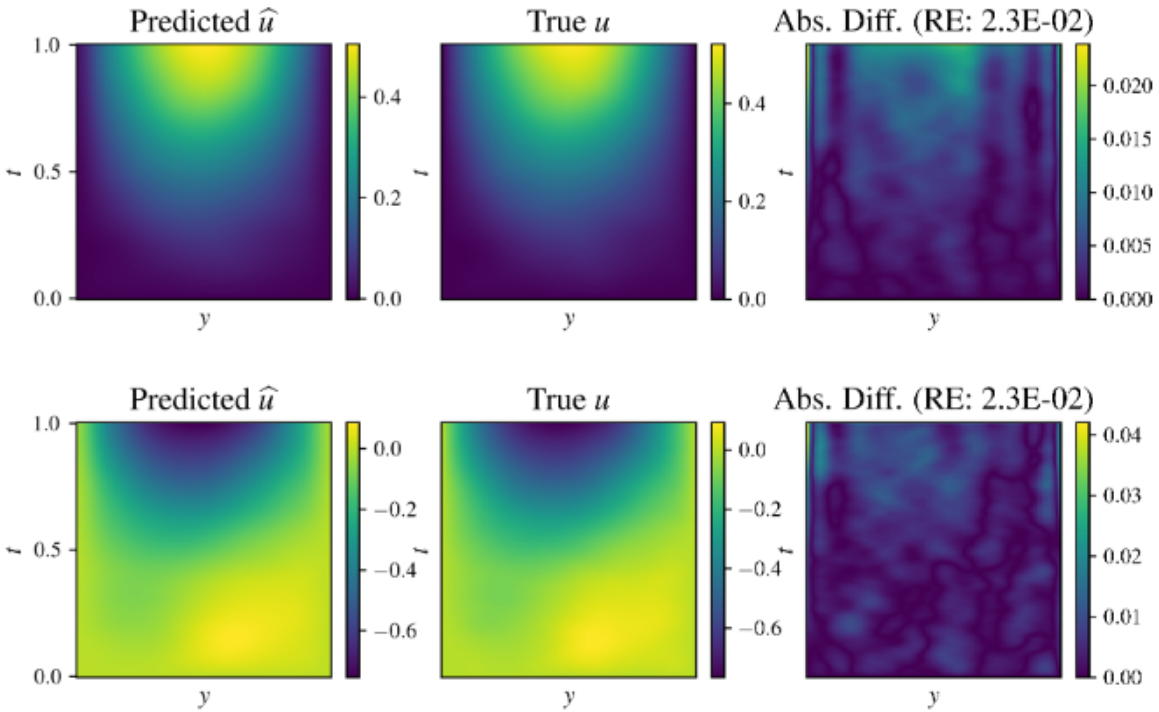
  where $\Omega \subseteq \mathbb{R}^n$ is some domain, f is a known function, and we have additional conditions on the boundary $\partial\Omega$.
- Risk function can be defined directly from the equation residual, or via input-output data; we may also consider operator learning $u = A(f)$.

Main reference: Stepaniants 2023

- In this article, the authors approximate the Green's function $G(x, y)$ for different PDEs of the form (1). Solutions u may then be written as:

$$u(x) = \int_{\Omega} G(x, y) f(y)\, dy.$$

- The following figure illustrates their final results for two $1D$ heat equation problems, $u_t - u_{xx} = f$ with Dirichlet boundary conditions:



Stepaniants 2023, cont.

- A few physics-informed constraints are considered, leading to improved results.
- For the $1D$ Helmholtz equation $\Delta u + \omega u = f$ $(\omega \in \mathbb{R})$, the Green's function has coordinate symmetry $G(x, y) = G(y, x)$, enforced by kernel transformation:

$$K_{sym}(x, y, x', y') = \frac{1}{4}\big[K(x, y, x', y') + K(x, y, y', x') + K(y, x, x', y') + K(y, x, y', x')\big].$$

- For the $1D$ heat equation, the Green's function satisfies time invariance $G(\cdot, s, \cdot, t) = G(\cdot, s - t, \cdot, 0)$, enforced by kernel transformation:

$$K_{inv}(x, s, y, t, \xi, \sigma, \eta, \tau) = K(x, y, t - s, \xi, \eta, \tau - \sigma).$$

- Hard to generalize, other constraints omitted.

My physics-informed approach (time-dependent problems)

- Idea: consider conservation law functionals for some density $q(u)$:

$$F[u](t) = \int_\Omega q(x,t)\,dx,$$

  that are upheld by solutions u to the PDE, i.e. $F[u]'(t) = 0$, and define RKHS that obeys this criteria.
- Found a lot in PDE literature, and in real models.
- Example: For the heat equation with Neumann boundary conditions and $f \equiv 0$, we have mass conservation:

$$\frac{d}{dt}\int_\Omega u(x,t)\,dx = \int_\Omega u_{xx}(x,t)\,dx = \int_{\partial\Omega} \frac{du}{d\mathbf{n}}\,dS = 0.$$

My physics-informed approach, cont.

- Example 2: for the $1D$ wave equation $u_{tt} - u_{xx} = 0$ with $u(0,t) = u(L,t) = 0$, the energy functional

$$E(t) = \int_{[0,L]} \frac{1}{2}u_t^2 + \frac{1}{2}u_x^2 \, dx,$$

  is conserved, i.e. $E'(t) = 0$.
- Example 3: for any $1D$ equation $u_t - \partial_x q(x,t,u,u_x,u_{xx},\ldots) = 0$ we have the (averaging principle) equality:

$$\int_{[0,L]} u(x,T) - u(x,0) \, dx = \int_{[0,T]} q(L,t) - q(0,t) \, dt.$$

- Captures global behavior of solutions. If incorporated into kernel design, the RKHS will automatically contain (partially) physics-validated functions.

How to define transformed RKHS

- Recall that our approximate functions in the RKHS are of the form:

$$\hat{u}(x,t) = \sum_{i,j} \alpha_{i,j} K(x,t,x_i,t_j),$$

  where $(x_i)_i$ and $(t_j)_j$ are basis-points. If the conservation law $F[u](t) = F[u](0)$ is linear in u, then we may equivalently enforce $F[K(\cdot,\cdot,x',t')](t) = F[K(\cdot,\cdot,x',t')](0)$ for all x' and t'.
- We consider a point-wise approach, to circumvent explicit representer computations needed for continuous centering.
- Let $\mathcal{T}_{con}$ be a set of time-points, and consider the system of equations $L_{t_a} = F[K](t_a) - F[K](0) = 0$ for all $t_a \in \mathcal{T}_{con}$.

How to define transformed RKHS, cont.

- Defining $L : \mathcal{H} \to \overbrace{\mathbb{R}^{N+1}}^{N+1=|\mathcal{T}_{con}|}$ by $f \mapsto (L_{t_0}[f], \ldots, L_{t_N}[f])^T$ and writing $Q(z') = L[K(\cdot, \cdot, z')]$ for $z' = (x', t')$ we show that the function:

$$\widetilde{K}(z, z') = K(z, z') - Q(z)^T A^{-1} Q(z'),$$

  where $z = (x, t)$ and $A = LL^*$ is the Gram matrix, satisfies the equation $L[\widetilde{K}(\cdot, z')] = 0$ for all z'.
- With finite background set $X = \mathcal{X}_{bas} \times \mathcal{T}_{bas}$ and a symmetric base kernel (i.e. Gaussian) we have that $A = Q(\mathcal{X}_{bas}, \mathcal{T}_{bas}) \cdot Q(\mathcal{X}_{bas}, \mathcal{T}_{bas})^T$.
- $\widetilde{K}$ is exactly the projection P of K into the null-space $\ker L$.
- But is $\widetilde{K}$ still a reproducing kernel? For what space?

Properties of transformed RKHS

- It is shown that if $K(\cdot, z') \in C^2(\Omega)$ for any z', then $\mathcal{H} \subseteq C^1(\Omega)$. In particular if K is infinitely differentiable, then so is any element of $\mathcal{H}$.
- This means that the operator L_{t_a} is bounded in $\mathcal{H}$ for all $t_a \in \mathcal{T}_{con}$, and in particular the projection into the combined null-space is a bounded linear operator (and ker$L$ is a closed subspace).
- The projection of K into ker$L$ has the reproducing property on ker$L$, since for any $f \in$ ker$L$:

$$\begin{aligned}\langle f, \widetilde{K}(\cdot, z)\rangle_{\text{ker}L} &= \langle f, P(K)(\cdot, z)\rangle_{\mathcal{H}} = \langle P^* f, K(\cdot, z)\rangle_{\mathcal{H}} \\ &= \langle f, K(\cdot, z)\rangle_{\mathcal{H}} = f(z),\end{aligned}$$

where we used that P is self-adjoint and acts as the identity on f.

Experiments

- We evaluate the transformed kernel method with two experiments, comparing to a baseline kernel expansion method.
- Experiment 1: $1D$ heat equation $u_t - u_{xx} = 0$ on the spatial domain $[-1, 1]$ with Neumann boundary conditions $u_x(-1, t) = u_x(1, t) = 0$ and initial condition u_0, using a residual based risk:

$$R(u) = \|u_t - u_{xx}\|^2_{L^2(\Omega)} + |u_x(-1, t)|^2 + |u_x(1, t)|^2 + \|u(x, 0) - u_0(x)\|^2_{L^2((-1,1))}$$

- The physics-informed constraint that we consider is the mass balance property:

$$L_{t_a}[u] := \int_{[-1,1]} u(x, t_a)\, dx - \int_{[-1,1]} u(x, 0)\, dx = 0, \quad \forall t_a \in \mathcal{T}_{con}.$$

Experiments, cont.

- Experiment 2: $1D$ nonlinear Richards equation:

$$\partial_t \theta - \partial_x (K(\theta) h'(\theta) \partial_x \theta + K(\theta)) = 0,$$

  where K and h are known functions, with Dirichlet boundary conditions. We use a FDM-based solver to compute reference data, and use a standard penalized loss.
- For this equation we consider the averaging principle. However, the flux q is nonlinear in u, so we construct a crude linearization procedure, and in the end consider:

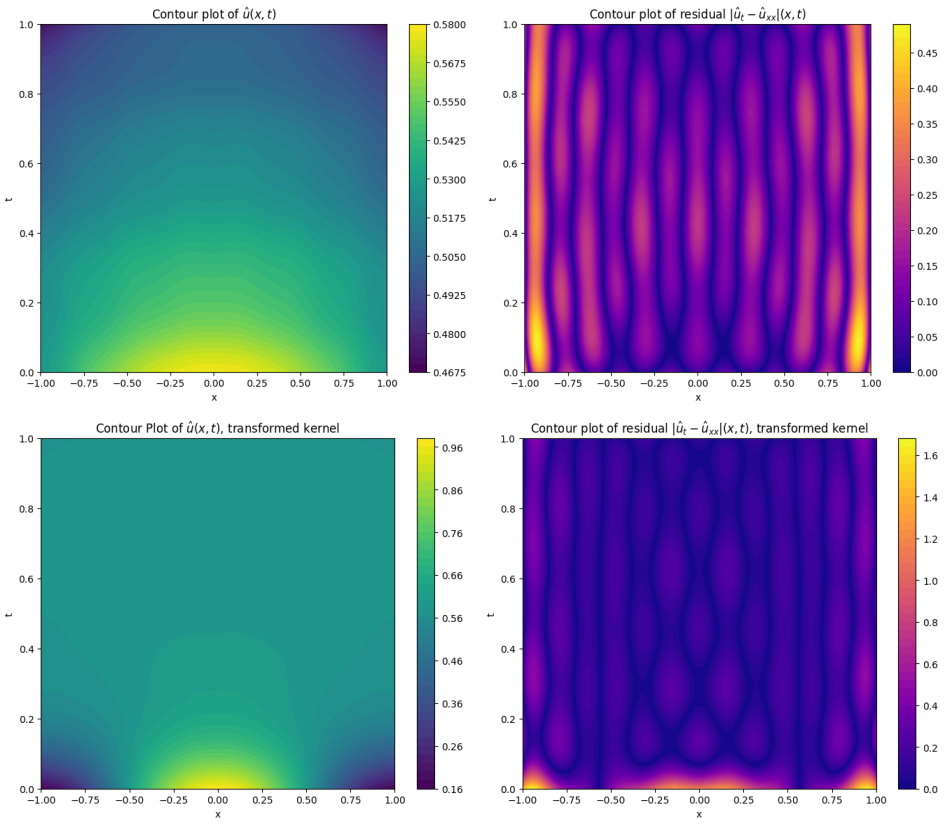
$$L_{t_a}[\theta] = \int_{[0,L]} \theta(x, t_a) - \theta(x, t_0)\, dx - \frac{1}{\alpha} \int_{[t_0, t_a]} \partial_x \theta(L, t) - \partial_x \theta(0, t)\, dt.$$

  for $\alpha \approx 113$.

Results for experiment 1

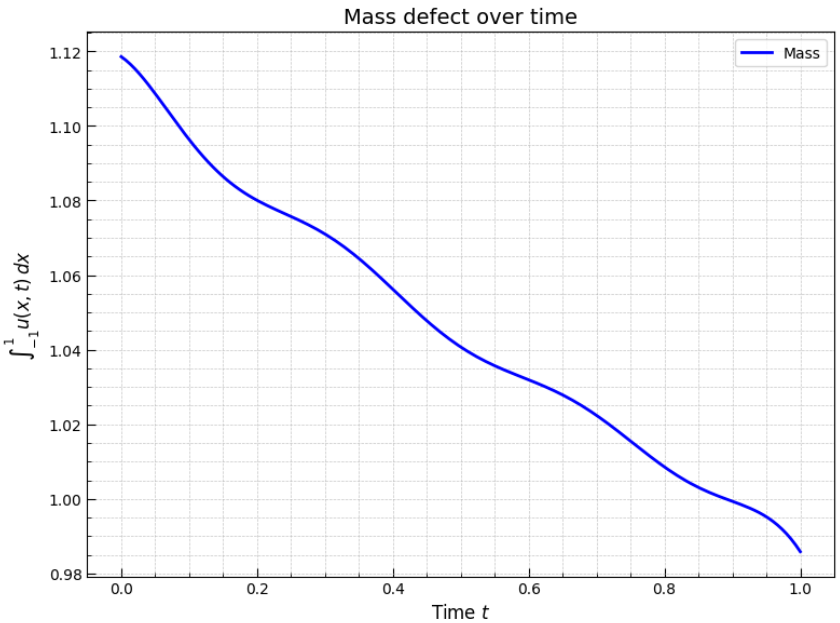
- We consider a "bumpy" initial condition $u_0(x) = \sin(\pi(x+1)/2)$, and compare results:

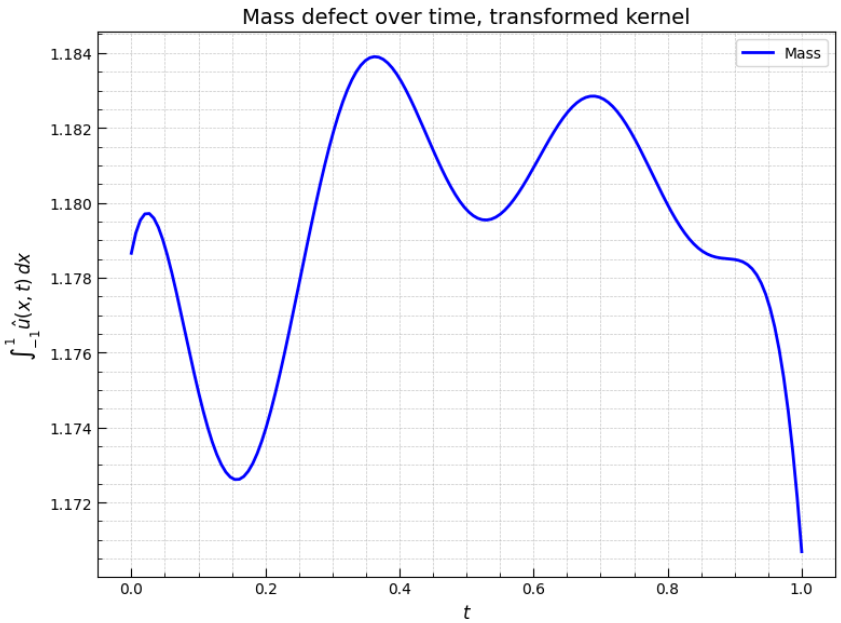
- The baseline converges to local (incorrect) minimizer.
- Improved residual, except at $t = 0$.
- Best accuracy is found with $N = 20$ basis points in each variable.
- Optimization time is comparable for this value, but is much larger in the transformed method for $N = 30$ and beyond.



Results for experiment 1, cont.

- We also see improvement in the mass conservation property, about two orders of magnitude.

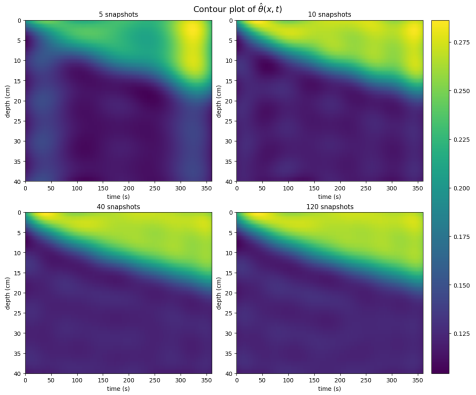


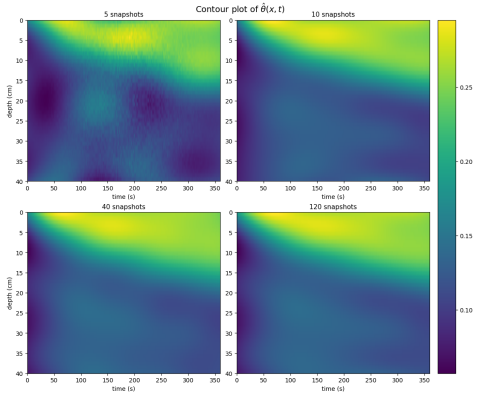


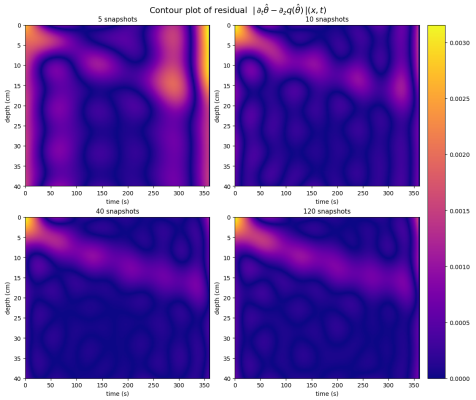
- Should be better, but high number of terms in kernel expansion, and the numerical error near $t = 0$ prevents this.

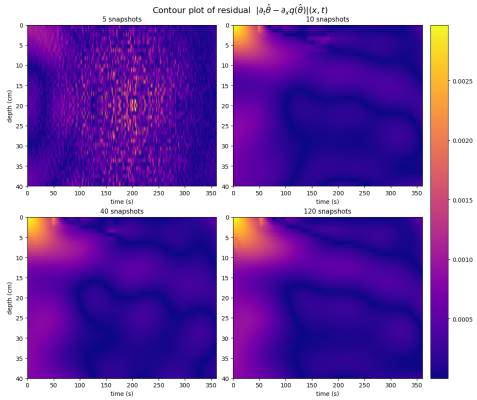
Results for experiment 2

- We see very marginal improvement for a constant inflow BVP (right column is the transformed kernel):





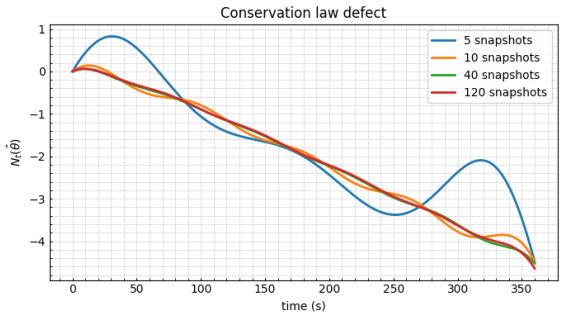


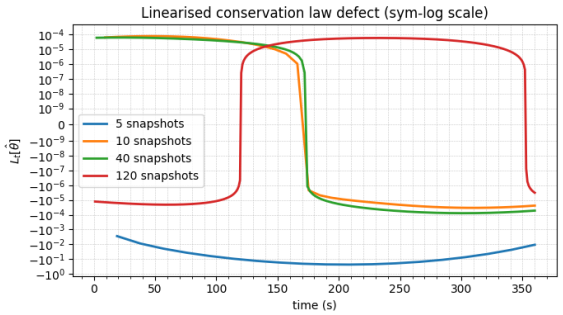


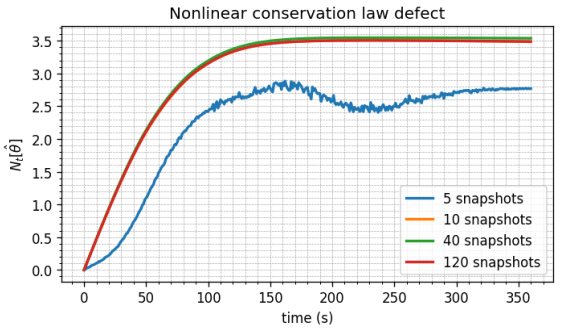
- Contour and residual plots are separated by how much reference data was used.
- Transformed kernel is more noise-sensitive.
- Structural residual errors are unchanged from the kernel transformation.

Results for experiment 2, cont.

- Even though the linearized conservation law is perfectly upheld:







- Still a large true conservation law error.
- Lack of improved results is likely due to the error from the linearized conservation law.
- A crude linear bound was derived, $|N_t[\theta] - L_t[\theta]| \leq Ct$ for some $C > 0$.
- Of limited practical use for large t.

Discussion

- Exact constraint yielded positive improvement, in particular residual-based regression only converged to true solution when additional physics was incorporated.
- Indicates that adding correct physics to kernel design is good for accuracy.
- Adding "incorrect" physics made little difference (neither positive or negative). Global constraint added to local bump.
- Increased computation time likely due to matrix multiplications needed to define transformed kernel. In theory, everything can be pre-computed so there should be no difference (ideally want something akin to Yu et al 2021).

Discussion, cont. (conclusion, future ideas)

- We presented methodology that allows for a RKHS kernel expansion to obey any linear (wrt. u)conservation law.
- Naive linearization of nonlinear conservation laws does not give positive results. Since many conservation laws are not linear, a method to better incorporate these are sought.
- Piecewise linear methods (unique data-obtained coefficients each timestep) is a possible solution.
- Toy models were used for experiments, but applications to real models can readily be constructed, and use-cases are increasing.
- In particular, a big improvement to results in Stepaniants is possible.

Thank you!