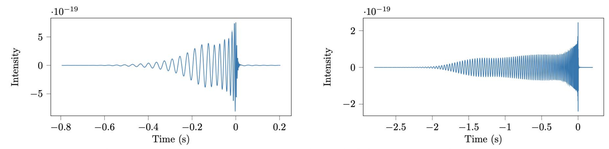
TpopT: Efficient Trainable Template Optimization on Low-Dimensional Manifolds

Jingkai Yan, Shiyu Wang, Xinyu Rain Wei, Zsuzsa Marka, Szabolcs Marka, John Wright

AI INSTITUTE IN DYNAMIC SYSTEMS

Introduction and Motivation

- Interesting signals/patterns often come from a low-dimensional manifold
  - Seismic data, MRI, natural images, etc
- Application: Gravitational wave (GW) detection (search for a parametric family of signals in noise)



  - Current method of choice: Matched filtering that uses a template bank with millions of templates – inefficient in high-dim parameter space!
- **Challenge**: How to detect signals/patterns from low-dimensional spaces more efficiently?
- Our approach: use **trainable optimization** instead of covering.

Problem Setup

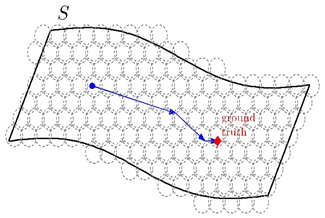
- **Task**: identify signal from input data **x**:

$$\boldsymbol{x} = \begin{cases} a\,\boldsymbol{s}_\natural + \boldsymbol{z} & \text{if contains signal } \boldsymbol{s}_\natural \in S \\ \boldsymbol{z}, & \text{if noise only} \end{cases}$$

- Matched filtering compares the input x with a template bank $\{\boldsymbol{s}_k\}_{k=1}^K$ and compute the max correlation
  - But for a given input, most templates are irrelevant!
- **Intuition**: it is more efficient to let optimization "guide" us to the best matching template
- Tpopt searches for a best matching template $\hat{\boldsymbol{s}}$, by numerically solving

$$\hat{\boldsymbol{s}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{s}\in S} f(\boldsymbol{s}) \equiv -\langle \boldsymbol{s}, \boldsymbol{x}\rangle .$$

through gradient descent $\boldsymbol{s}^{k+1} = \exp_{\boldsymbol{s}^k}\left(-\tau_k \operatorname{grad}[f](\boldsymbol{s}^k)\right)$.



Main Contributions

- Theoretically proved the efficiency gain: exponential in dim(S), larger gain when required accuracy is higher
- Proposed a trainable network architecture constructed via unrolled optimization
- Achieved better dimension scaling for signal detection problems – better able to search higher dimensional space and discover interesting phenomena

Theory: Efficiency gains over Matched Filtering

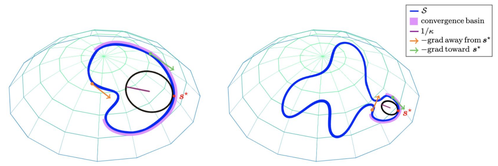
Theorem 1. *Let S be a complete manifold. Suppose the extrinsic geodesic curvature of S is bounded by κ. Consider the Riemannian gradient method, with initialization satisfying $d(\boldsymbol{s}^0, \boldsymbol{s}_\natural) < 1/\kappa$, and step size $\tau = \frac{1}{64}$. Then when $\sigma \le c/(\kappa\sqrt{d\log D})$, with high probability, we have for all k*

$$d\left(\boldsymbol{s}^{k+1}, \boldsymbol{s}_\natural\right) \le (1-\epsilon)\, d\left(\boldsymbol{s}^{k}, \boldsymbol{s}_\natural\right) + C\sigma\sqrt{d\log(\kappa D)}.$$

Moreover, when $\sigma \le c/(\kappa\sqrt{D})$, with high probability, we have for all k

$$d(\boldsymbol{s}^k, \boldsymbol{s}^\star) \le (1-\epsilon)^k d(\boldsymbol{s}^0, \boldsymbol{s}^\star),$$

where $\boldsymbol{s}^\star$ is the unique minimizer of f over $B(\boldsymbol{s}_\natural, 1/\kappa)$. Here, C, c, ϵ are positive numerical constants.



Convergence to **Optimal Statistical Precision**

- Linear convergence to a $\sigma\sqrt{d\log D}$- neighborhood of $\boldsymbol{s}_\natural$ which is within a logarithmic factor of optimal.
- Accuracy and the noise level are dictated by the intrinsic dimension d, with only a logarithmic dependence on the ambient dimension.
- Under a stronger condition on σ – linear convergence for all iterations

The **complexities** of the two methods - MF and TpopT - are given by

$$T_{\mathrm{MF}} \propto \frac{1}{r^d}, \qquad T_{\mathrm{TpopT}} \propto \kappa^d \log\frac{1}{\kappa r}.$$

where r is the target estimation accuracy

- TpopT always has superior dimensional scaling than MF whenever the allowable estimation error r is below 1/κ (and identical to MF above that). The advantage is more significant at lower noise and higher estimation accuracy.

Method : An efficient trainable architecture

Overall idea:

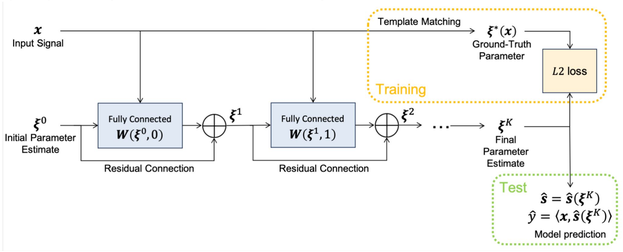
- First formulate an optimization problem and write down a GD solver
- Then re-formulate its components (Jacobians, step sizes, smoothing levels) to become trainable - each GD iteration is one layer of the network

The GD solver:

- Create a low-dim embedding of signals: $\boldsymbol{s} \mapsto \boldsymbol{\xi}$
- GD on the embedding space $\boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k + \underset{\text{step size}}{\tau_k} \cdot [\underset{\text{Jacobian}}{\nabla_{\boldsymbol{\xi}} s(\boldsymbol{\xi}^k)}]^T \boldsymbol{x}$
- Use kernel interpolation to estimate the Jacobians
- Use multi-level smoothing to facilitate optimization

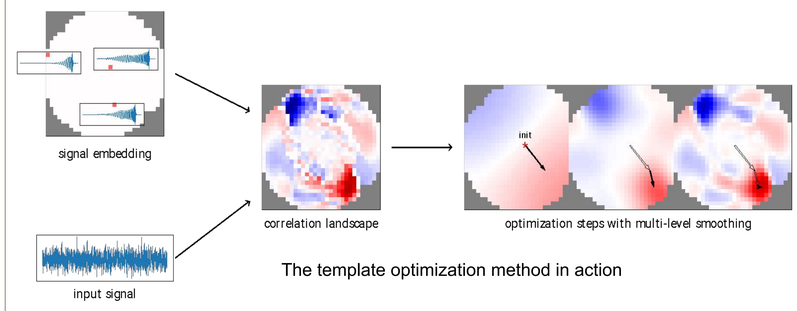
Re-formulate GD to be trainable: $\boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k + \underset{\text{trainable weights}}{\boldsymbol{W}(\boldsymbol{\xi}^k, k)}\, \boldsymbol{x}$

- $\boldsymbol{W}(\boldsymbol{\xi}^k, k)$ interpolated from a family of trainable "base" W matrices



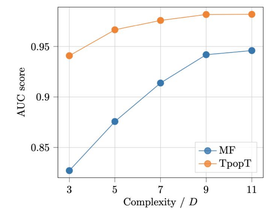
Results and Discussion

- Application: gravitational wave detection

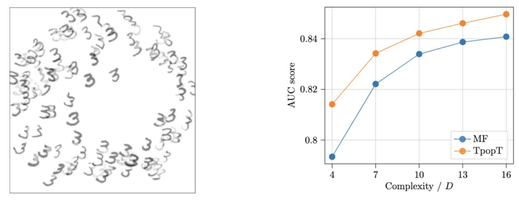


The template optimization method in action

- Comparison of accuracy-complexity curves for GW detection



- Application: handwritten digit recognition



Left: A slice of the 3-d embeddings projected onto the first two dimensions. **Right:** Classification scores of MF and TpopT at different complexity levels, for handwritten digit recognition.

- In both applications, TpopT achieves consistent higher detection accuracy compared with MF at equal complexity

Future Work

- Model discovery - learn aspects of signal manifold from data
- Convolutional extension - enjoy both the dimension-scaling advantage of TpopT and the signal-length scaling of matched filtering
- Applications to other kinds of dynamic data (e.g., EKG signals, data from imaging / materials)

References

Jingkai Yan, Shiyu Wang, Xinyu Rain Wei, Zsuzsa Marka, Szabolcs Marka, and John Wright. Tpopt: Efficient trainable template optimization on low-dimensional manifolds. preprint, 2023.

Jingkai Yan, Mariam Avagyan, Robert E Colgan, Do ̃ga Veske, Imre Bartos, John Wright, Zsuzsa Márka, and Szabolcs Márka. Generalized approach to matched filtering using neural networks. Physical Review D, 105(4):043006, 2022

Jingkai Yan, Robert Colgan, John Wright, Zsuzsa Márka, Imre Bartos, and Szabolcs Márka. Boosting the efficiency of parametric detection with hierarchical neural networks. Physical Review D, 106(6):063008, 2022.