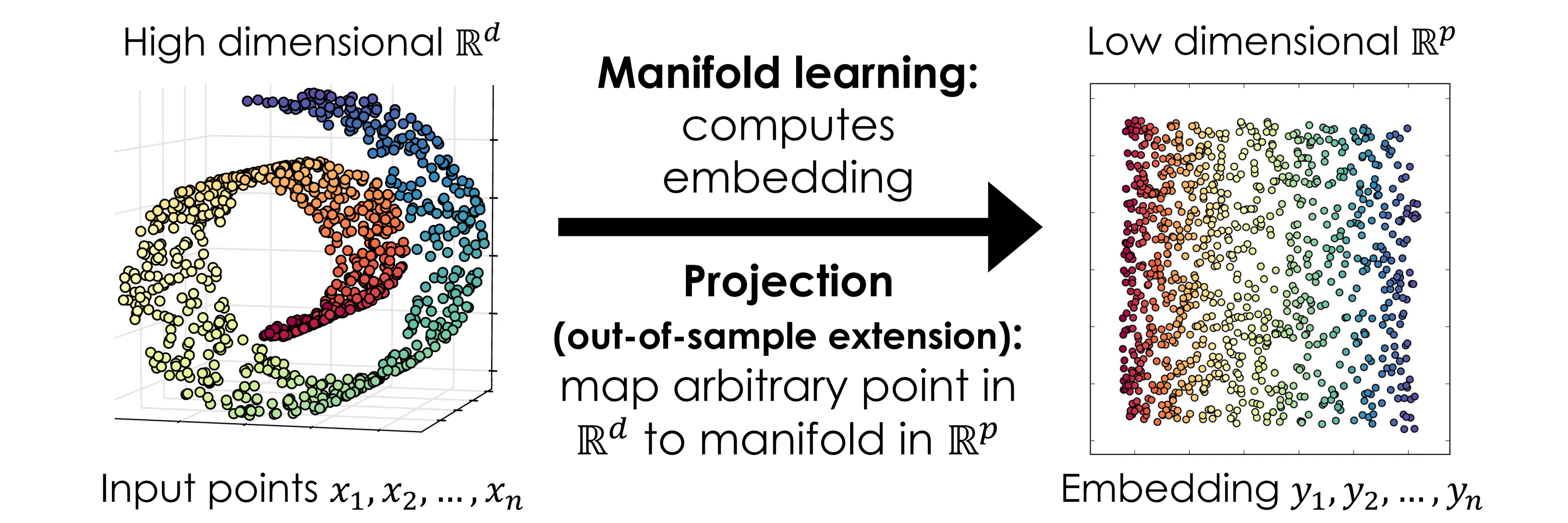
Sparse Projections of Medical Images onto Manifolds

George H. Chen, Christian Wachinger, Polina Golland

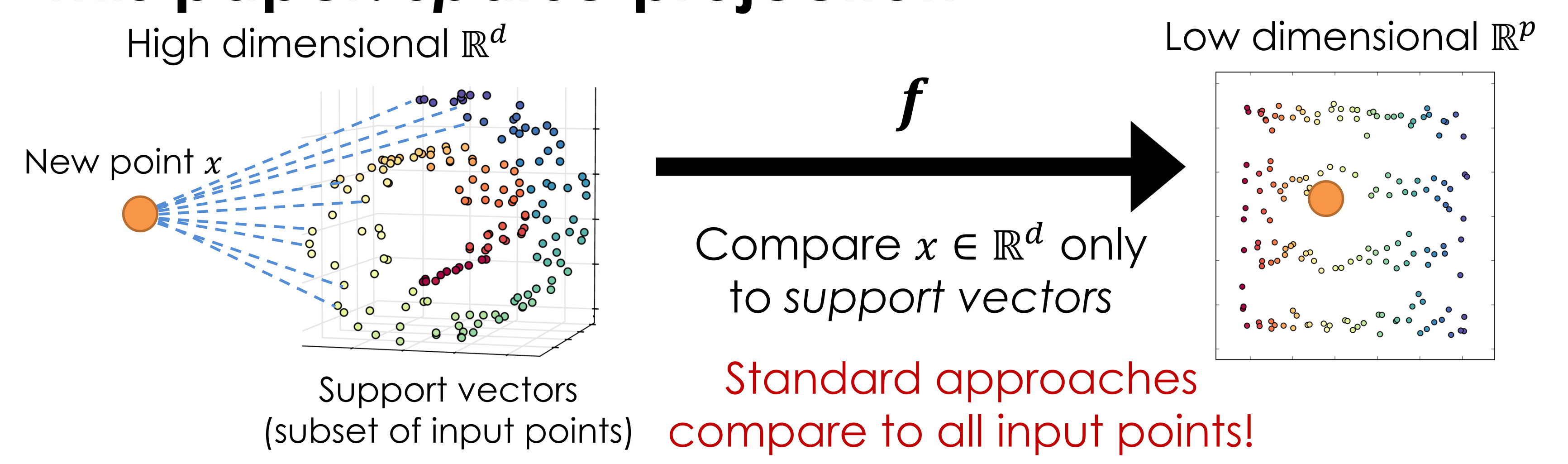
Motivation

Manifold learning in medical imaging
→ *e.g., segmentation, registration, computational anatomy, classification, detection, respiratory gating*

Some applications demand fast projection to manifold



This paper: *sparse* projection



Main idea

- Use any manifold learning algorithm to embed input points $x_1, x_2, \dots, x_n \in \mathbb{R}^d$ to $y_1, y_2, \dots, y_n \in \mathbb{R}^p$ $(d \gg p)$
- Compute "good" projection f that maps $\mathbb{R}^d$ to $\mathbb{R}^p$ and depends on only a few of $x_1, x_2, \dots, x_n$ (*support vectors*)
- Computational cost proportional to # support vectors!
- Trade off projection accuracy with computational cost

Sparse Kernel Ridge Regression

$\mathbb{H}$: reproducing kernel Hilbert space of functions $\mathbb{R}^d \to \mathbb{R}^p$
$\mathbb{K}(x, x')$: kernel for $\mathbb{H}$ specifying how similar $x, x' \in \mathbb{R}^d$ are

Kernel ridge regression

$$\hat{f} = \arg\min_{f \in \mathbb{H}} \Big\{ \underbrace{\sum_{i=1}^{n} \|y_i - f(x_i)\|_2^2}_{\text{Data fit}} + \underbrace{\lambda \|f\|_{\mathbb{H}}^2}_{\text{Smoothness}} \Big\}$$

Solution: $\hat{f}(\cdot) = \sum_{i=1}^{n} \mathbb{K}(\cdot, x_i)\hat{\alpha}_i$, $\quad \hat{\alpha} = (K + \lambda I_{n \times n})^{-1} Y$, $\quad K_{ij} = \mathbb{K}(x_i, x_j)$

Depends on *all* input points!

Sparse kernel ridge regression

Compute $\tilde{f}(\cdot) = \sum_{i=1}^{n} \mathbb{K}(\cdot, x_i)\tilde{\alpha}_i$ where
- Many $\tilde{\alpha}_i$'s are zero
- $\frac{1}{n}\sum_{i=1}^{n} \|\hat{f}(x_i) - \tilde{f}(x_i)\|_2^2 \le \varepsilon^2$

guarantee ε approx. to KRR (higher ε → fewer support vectors)

Depends only on support vectors (x_i for which $\tilde{\alpha}_i \neq 0$)

Formulate as convex program:

$$\tilde{\alpha} = \arg\min_{\alpha \in \mathbb{R}^{n \times p}} \sum_{i=1}^{n} \|\alpha_i\|_2 \quad \text{s.t.} \quad \|K\hat{\alpha} - K\alpha\|_F^2 \le n\varepsilon^2$$

mixed ℓ_1/ℓ_2 norm encourages many $\tilde{\alpha}_i$'s to be 0

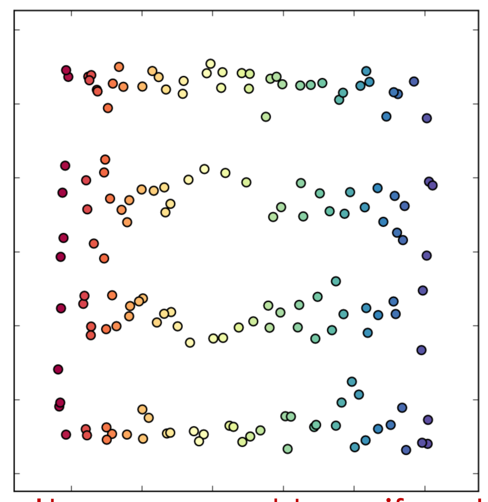
→ *solve with FISTA*

Results

Synthetic data

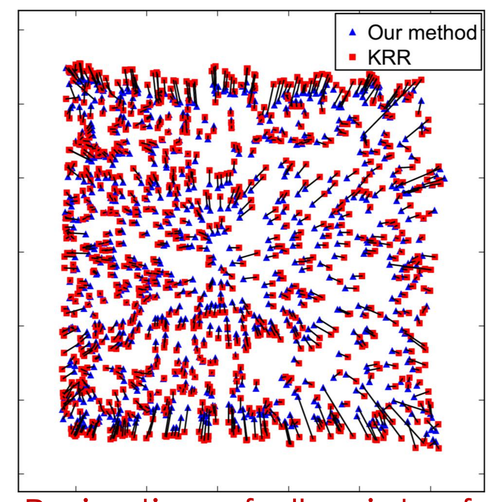
- 1000-pt Swiss roll → 2D (Hessian eigenmaps)
- Kernel: $\mathbb{K}(x, x') = \exp(-\|x - x'\|_2^2 / \sigma^2)$

Projected support vectors (λ=0.1, σ=4, ε=0.003)



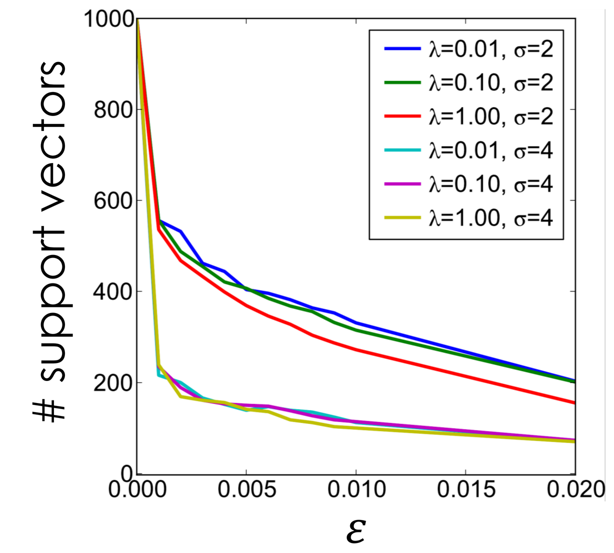
Don't correspond to uniformly sampled points in input space!

KRR vs. our method (λ=0.1, σ=4, ε=0.003)



Projection of all points of Swiss roll

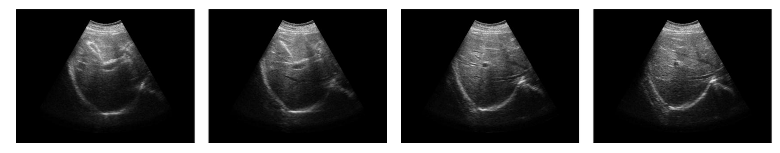
Higher λ or σ result in fewer support vectors



 # support vectors seems to scale with "complexity" of embedding, not number of training data points (λ=0.1, σ=4, ε=0.003)

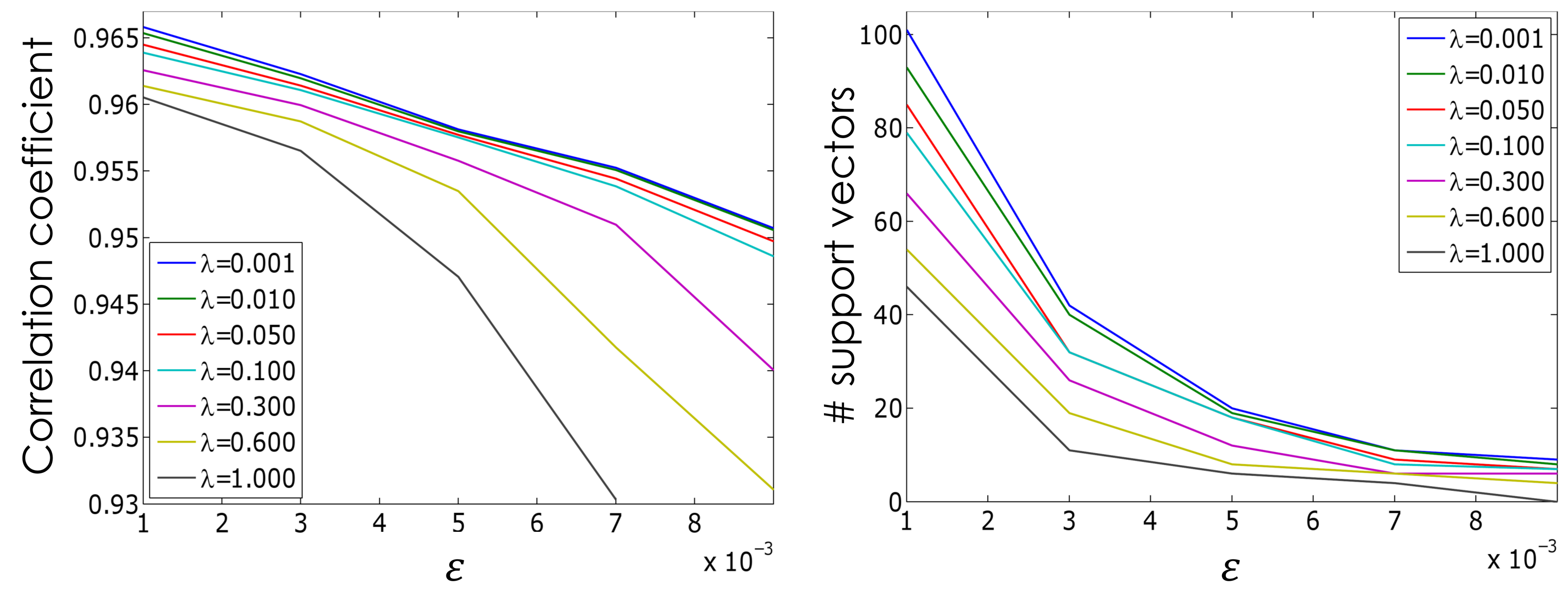
# pts in Swiss roll	1000	2000	3000	4000
# support vectors	161	174	163	170

Respiratory gating for ultrasound



- 640x480 images → 1D (Laplacian eigenmaps)
- 1D manifold enables tracking patient breathing cycle
- Train on first 200 frames of sequence, embed rest using sparse projection, correlate with embedding of full seq. (repeat with 5 seq's of lengths 354, 335, 298, 371, 298)

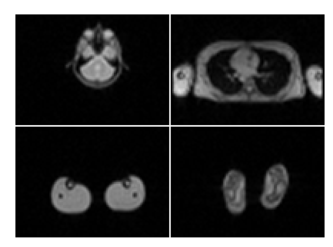
Trade off correlation coefficient & computational complexity



 # support vectors similar between embedding of first 200 frames and embedding of full seq. (λ=0.1, ε=0.001)

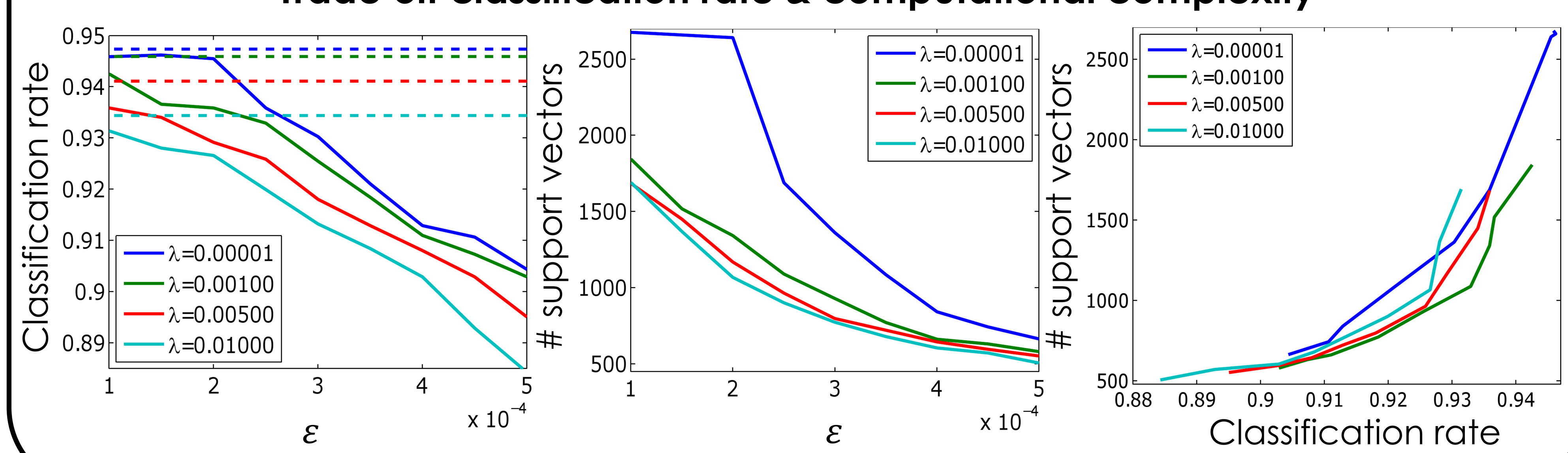
		Seq 1	Seq 2	Seq 3	Seq 4	Seq 5
# support vectors	First 200 frames	79	99	51	53	41
	Full seq.	73	100	61	45	50

MRI classification to monitor tissue heating



- 64x64 axial images → 2D (Laplacian eigenmaps)
- Nearest-neighbor classifier in 2D manifold labels each image as belonging to a body part (head, neck, lung, etc.)
  → real-time estimation of patient position in scanner

Trade off classification rate & computational complexity



This work is supported by: NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NIBIB NAC P41-EB-015902, NDSEG Fellowship (GHC)