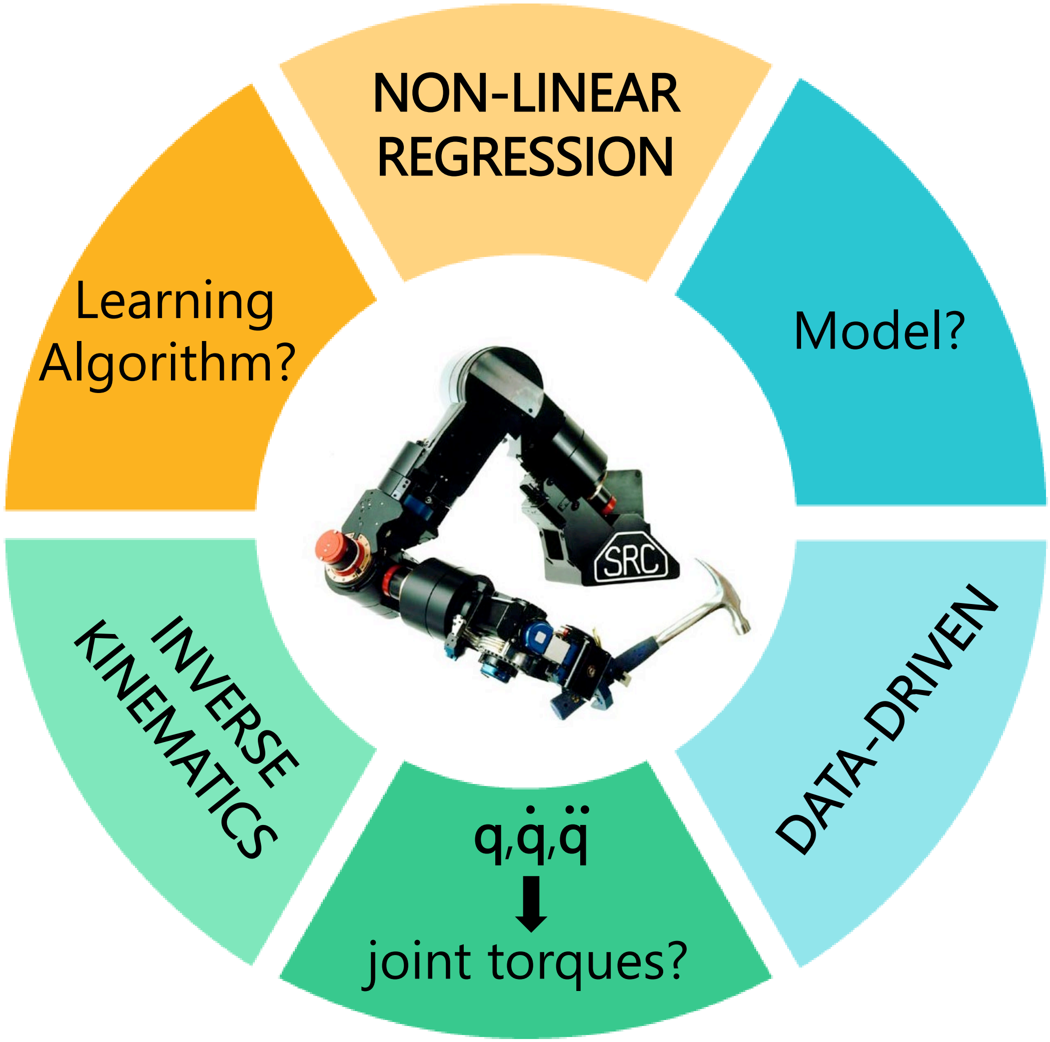
SCALING UP GAUSSIAN PROCESSES WITH TENSOR-BASED METHODS

CLARA MENZEN

Manon Kok, Kim Batselier



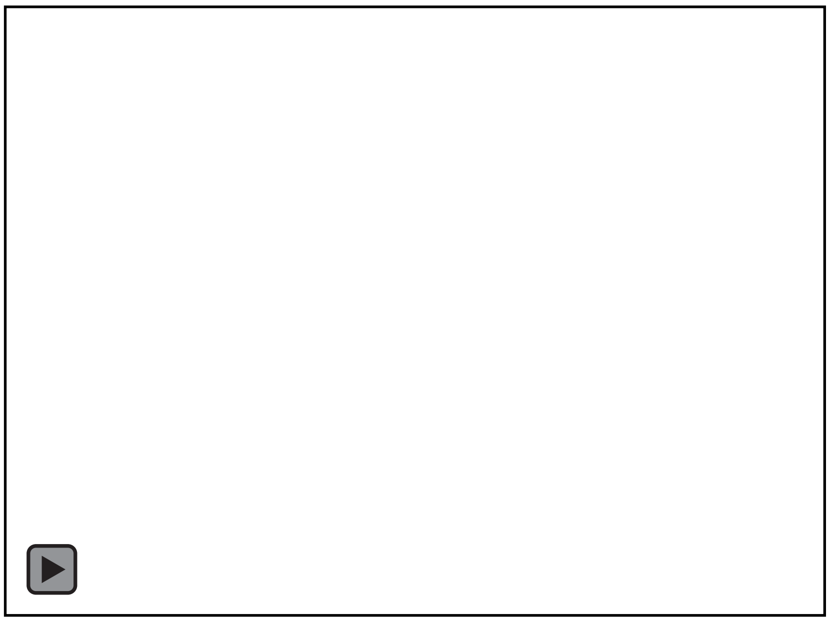
[Vijayakumar & Schaal 2000]

NON-LINEAR REGRESSION

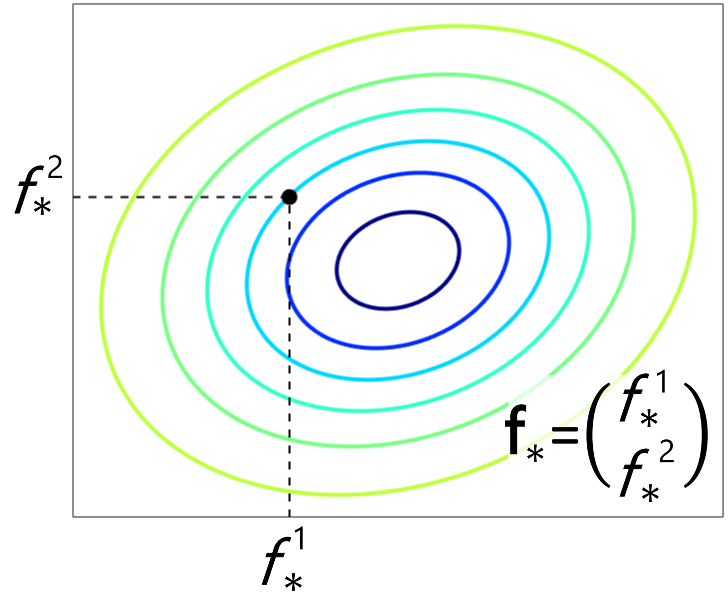
(play video)

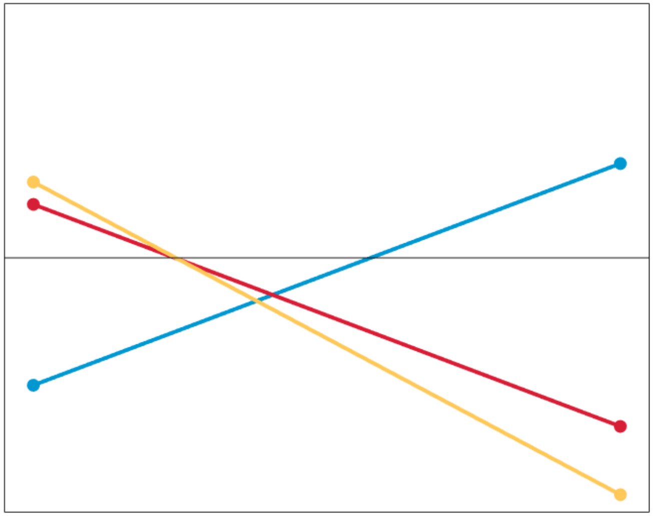
[Henning 2013]

GAUSSIAN PROCESSES

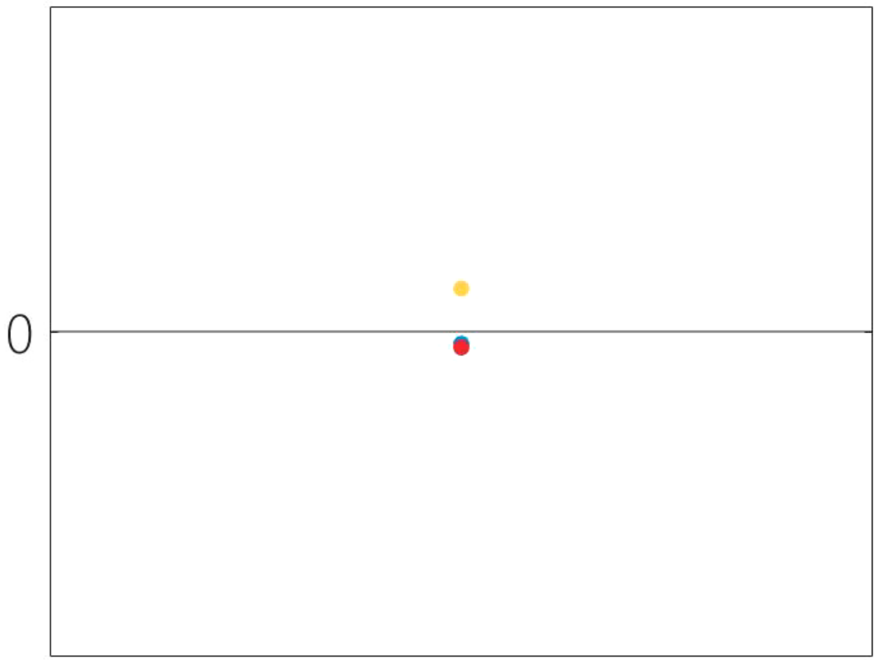


Bivariate Gaussian
(play video)





GAUSSIAN PROCESSES



(play video)

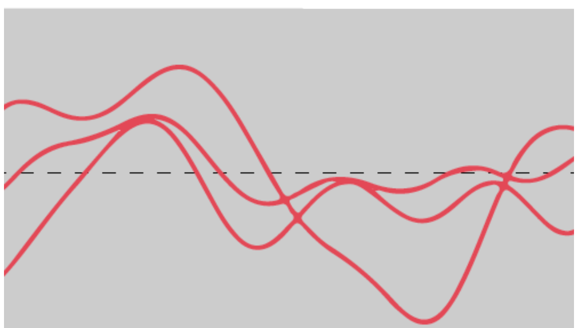
Distribution over functions

Fully specified by its
mean and covariance function

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}'))$$

GAUSSIAN PROCESS REGRESSION

Prior Distribution



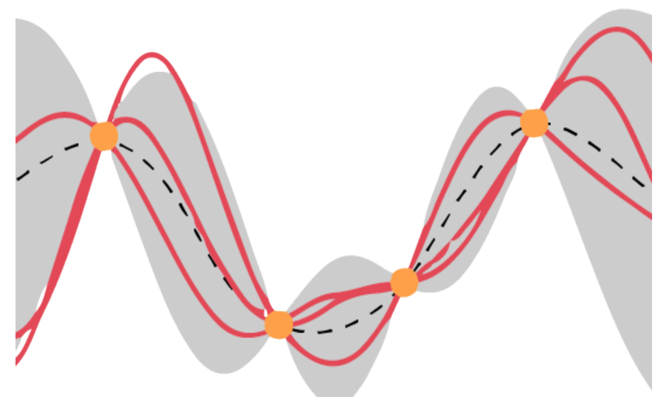
$$\mathbf{f}_* \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{**})$$

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

$$\mathrm{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix}\right)$$

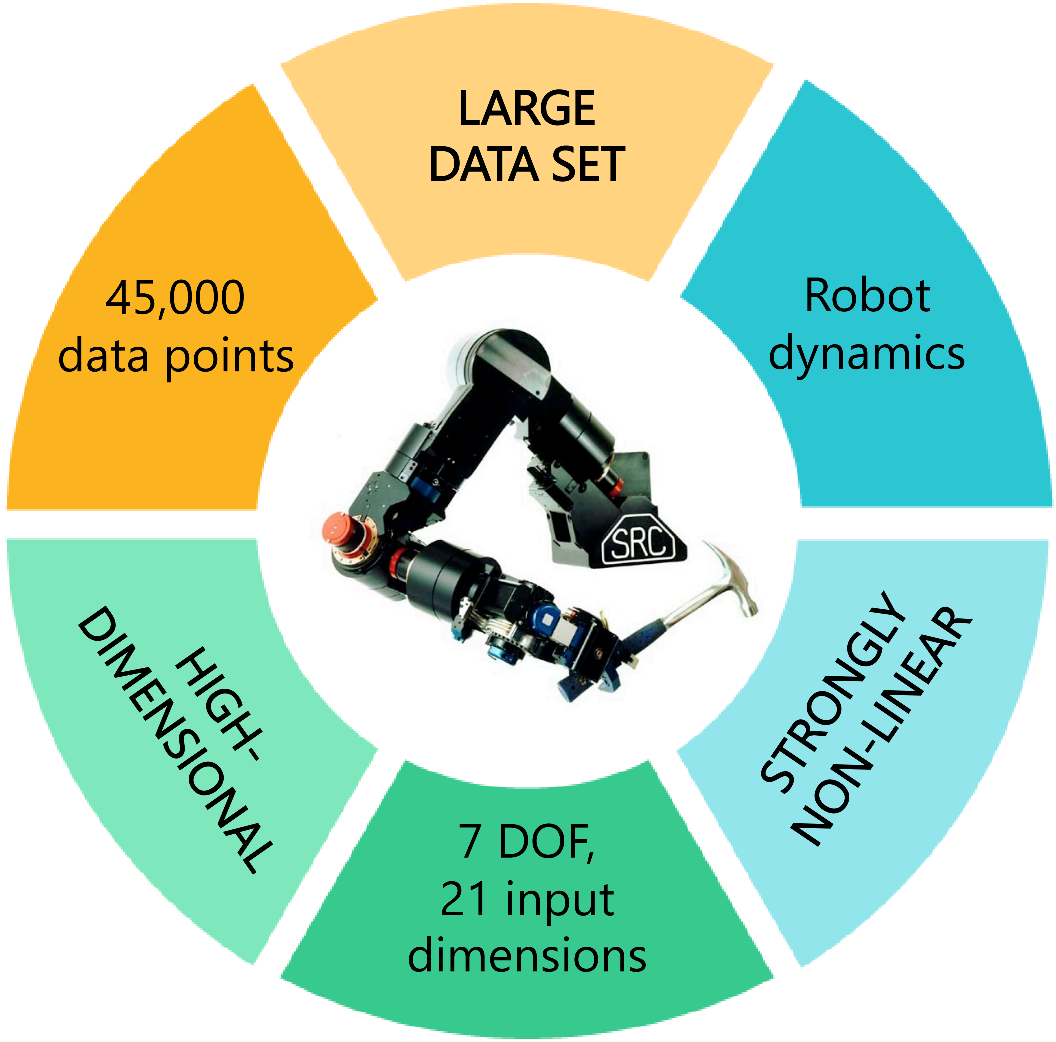
Posterior Distribution



$$\mathbf{f}_* | \mathbf{x}_*, \mathbf{x}, \mathbf{y} \sim \mathcal{N}(\mathbf{m}_{post}, \mathbf{K}_{post})$$

$$\mathbf{m}_{post} = \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{y}$$

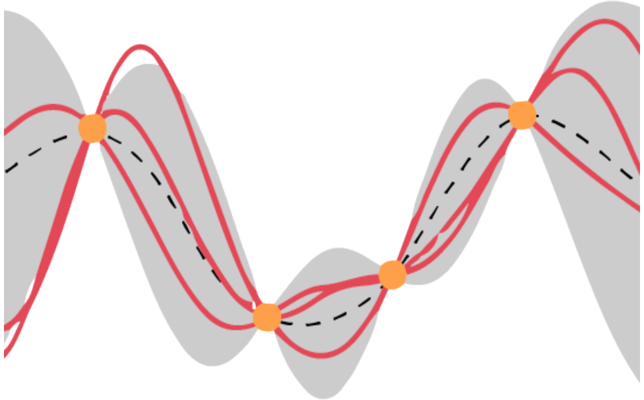
$$\mathbf{K}_{post} = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*$$



[Vijayakumar & Schaal 2000]

GAUSSIAN PROCESS REGRESSION

Posterior Distribution



$$\mathbf{f}_* | \mathbf{x}_*, \mathbf{x}, \mathbf{y} \sim \mathcal{N}(\mathbf{m}_{post}, \mathbf{K}_{post})$$

$$\mathbf{m}_{post} = \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{f}$$

$$\mathbf{K}_{post} = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*$$

K

Entries computed with covariance function

Size: N x N (N: number of data points)

Inverse needed: $O(N^3)$

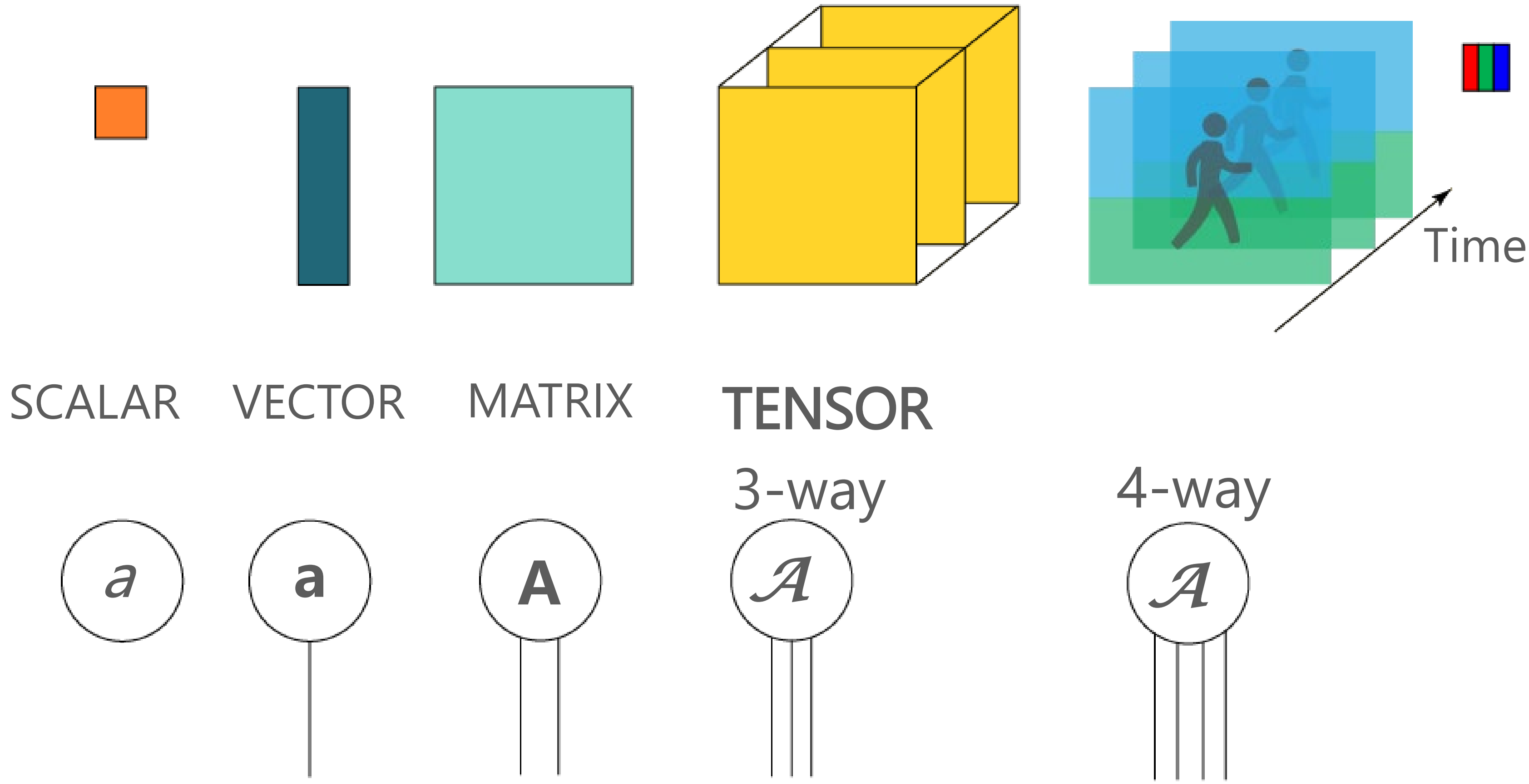
$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

$$\mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$$

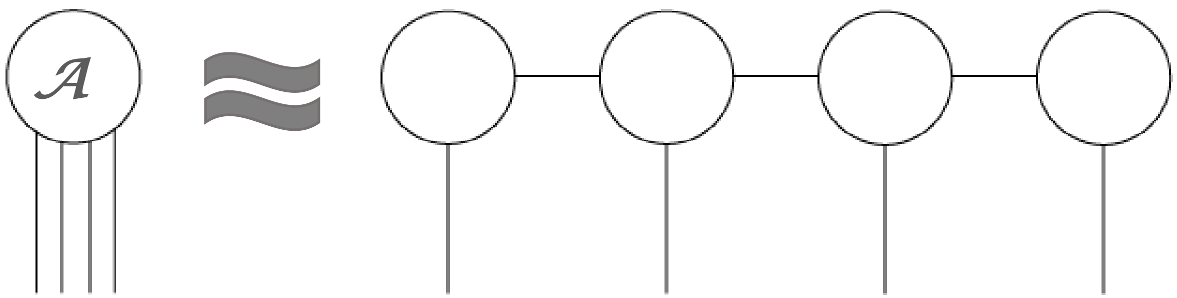
TENSOR METHODS

HIGH PERFORMANCE COMPUTING

INDUCING INPUTS


Time
SCALAR
VECTOR
MATRIX
TENSOR
3-way
4-way
a
a
A
𝒜
𝒜

TENSOR decom-positions

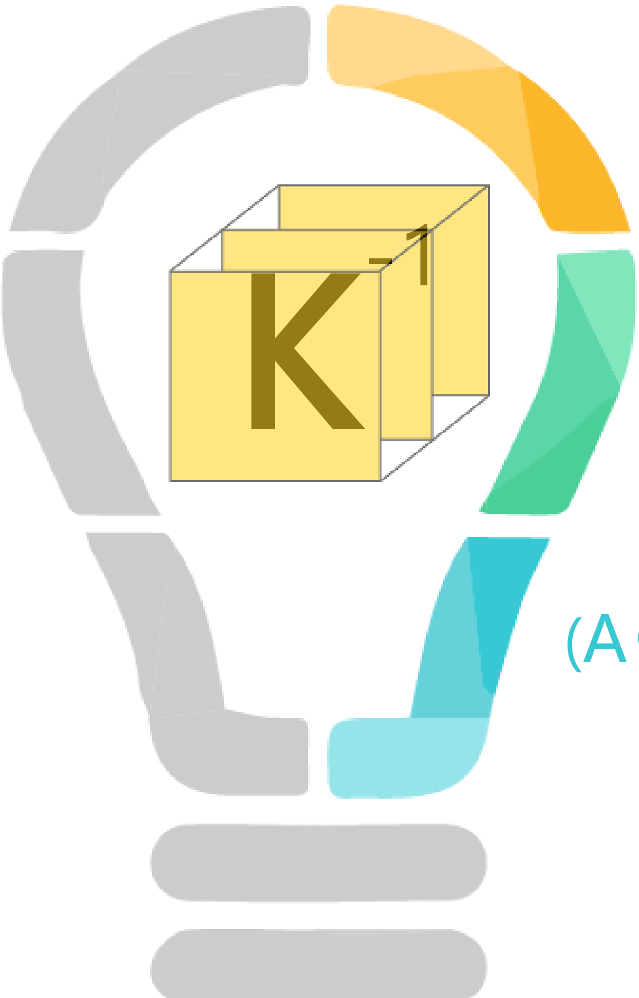


Tensor Train Decomposition

[Oseledets 2011]

TENSOR decom- positions

(play video)



$k\,(x,x')$ New covariance function

$K\,K^{-1} = I$ Alternating Linear Scheme

$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ Tensor Algebra

[Dolgov & Savostyanov 2014]
[Saatçi 2011]

Sources

[1] S. Vijayakumar & S. Schaal: Locally Weighted Projection Regression: An O(n) Algorithm for Incremental Real Time Learning in High Dimensional Space. Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000). Vol. 1., 2000

[2] P. Henning: Animating Samples from Gaussian Distributions, Technical Report No. 8, of the Max Planck Institute for Intelligent Systems, 2013

[3] C.E. Rasmussen: Gaussian Processes for Machine Learning, MIT Press, 2006

[4] I.V. Oseledets: Tensor-Train Decomposition, SIAM J. Sci. Comput. 33(5): 2295-2317, 2011

[5] Y. Saatçi: Scalable Inference for Structured Gaussian Process Models, PhD thesis, 2011

(play video)