Fluid Dynamics Models for Low Rank Discriminant Analysis

Yung-Kyun Noh

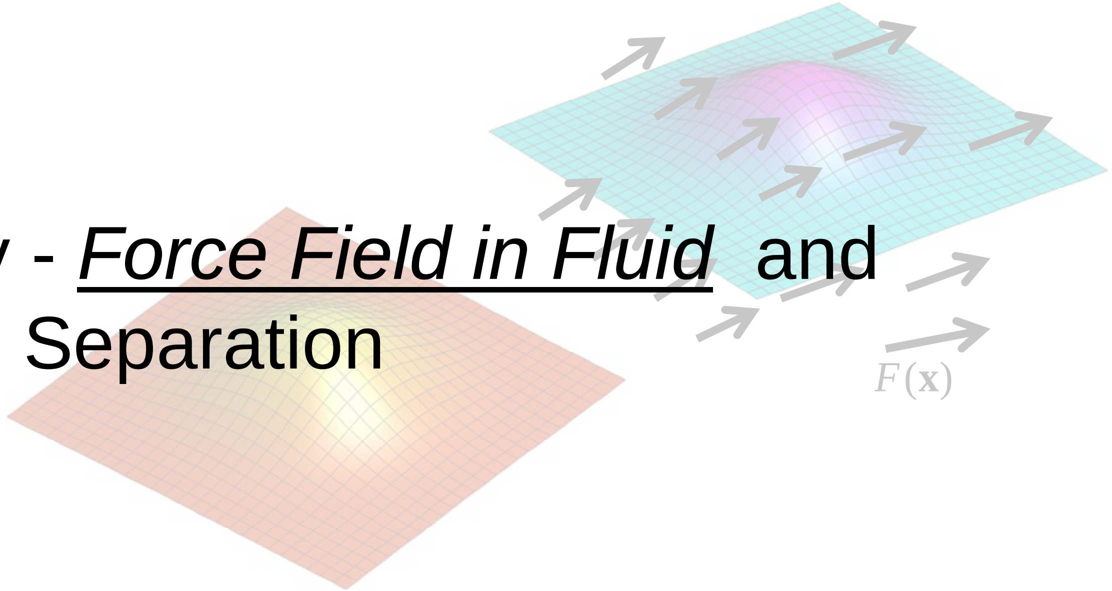
Byoung-Tak Zhang

Daniel D. Lee

AI & Statistics 2010

Overview

- Dimensionality Reduction Problem for Classification – Bhattacharyya Analysis
- Physics Analogy - ***Force Field in Fluid*** and Motion for Class Separation



- Discrimination of Gaussian Processes

Discriminant Analysis

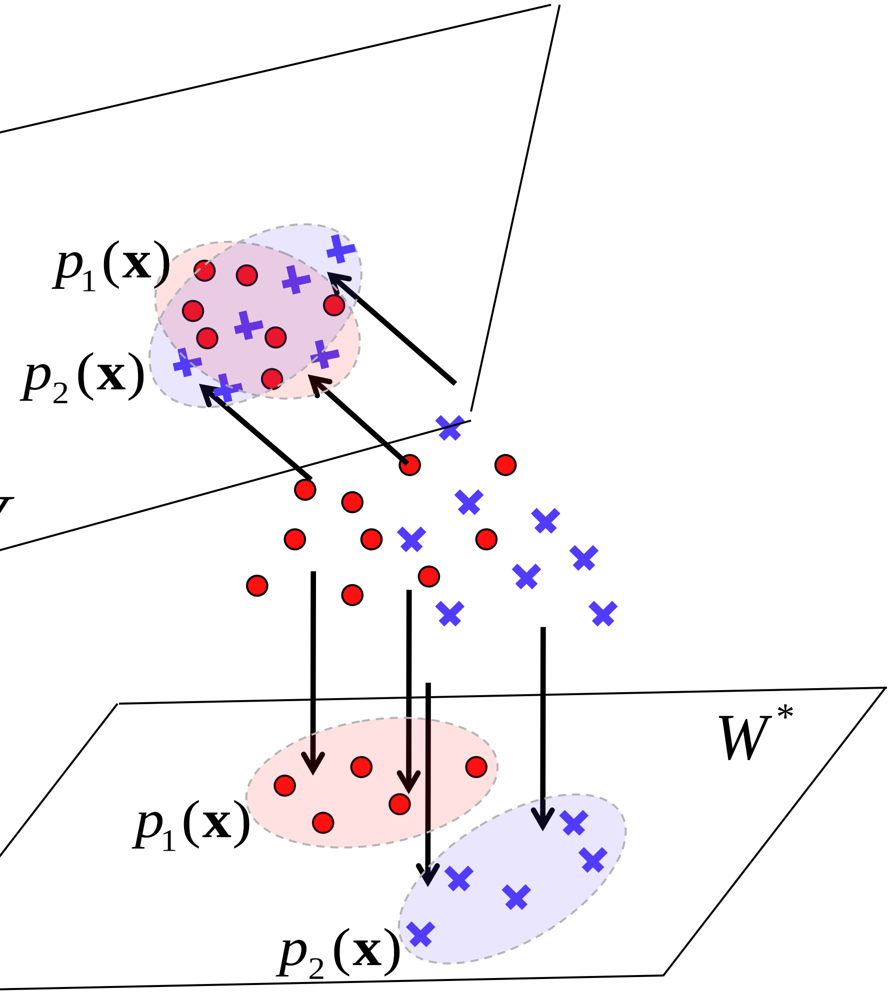
- Maximize class separation for classification.

Optimal subspace

$$W^* = \arg\min_{W} E_{Bayes}(W)$$

Objective function

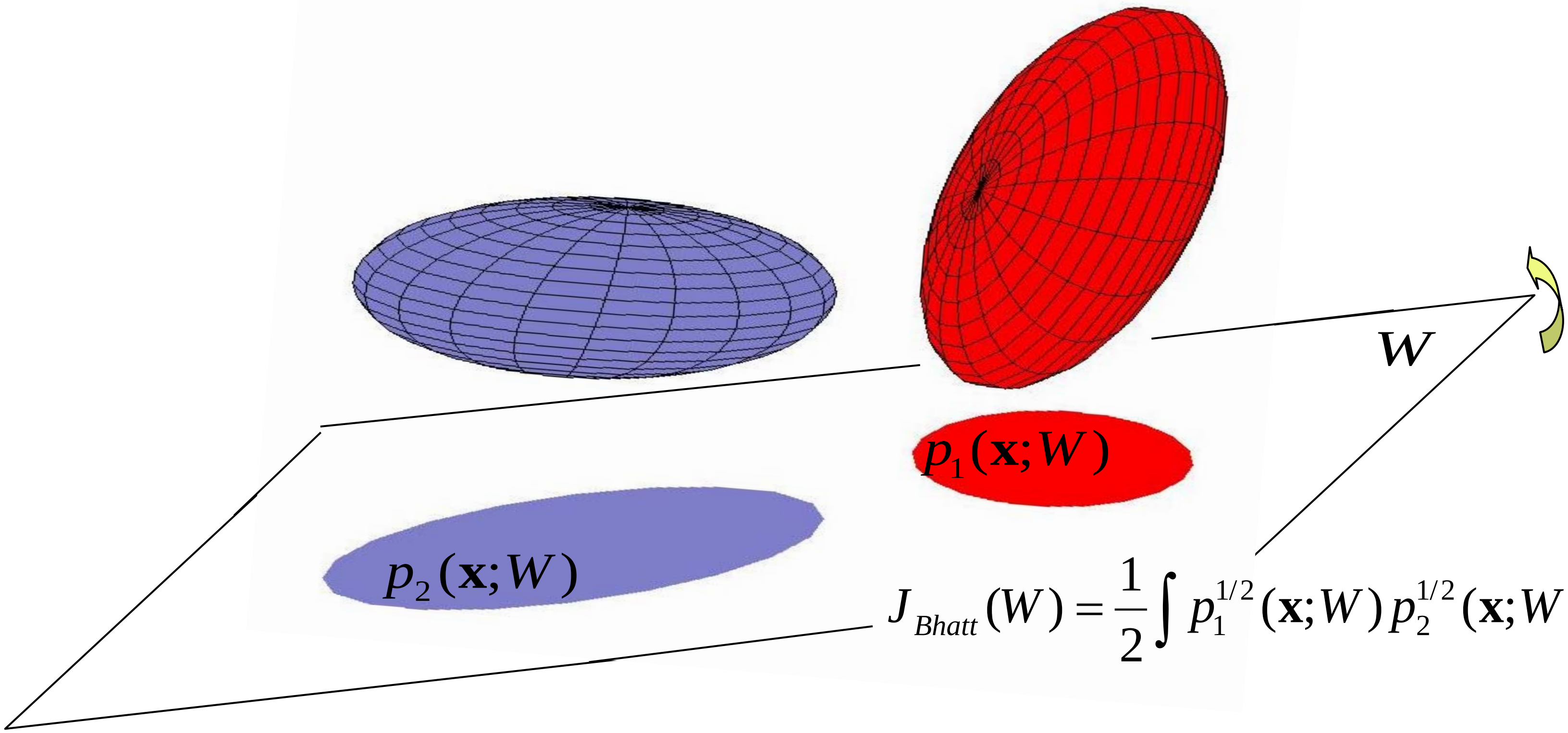
$$E_{Bayes} = \frac{1}{2}\int \min[p_1(\mathbf{x}), p_2(\mathbf{x})]\,d\mathbf{x}$$



Analysis with Bhattacharyya Bound

- Bhattacharyya bound with Gaussian distributions is integrable.

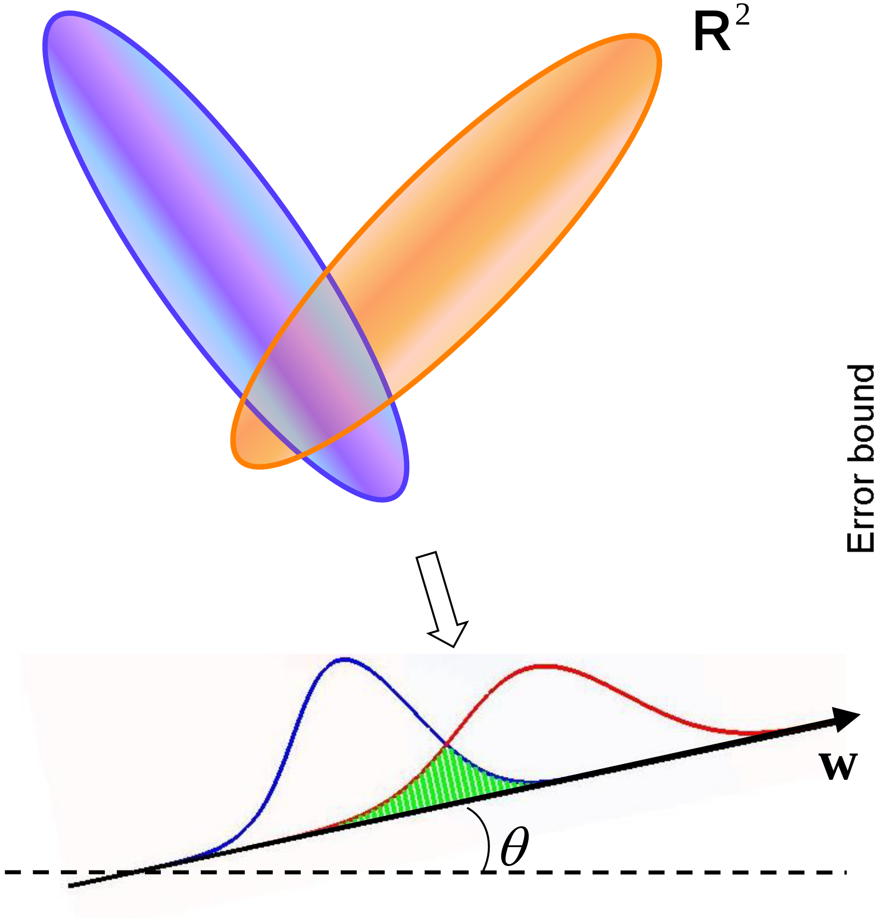
$$\min[p_1, p_2] \le p_1^{1/2} p_2^{1/2}$$



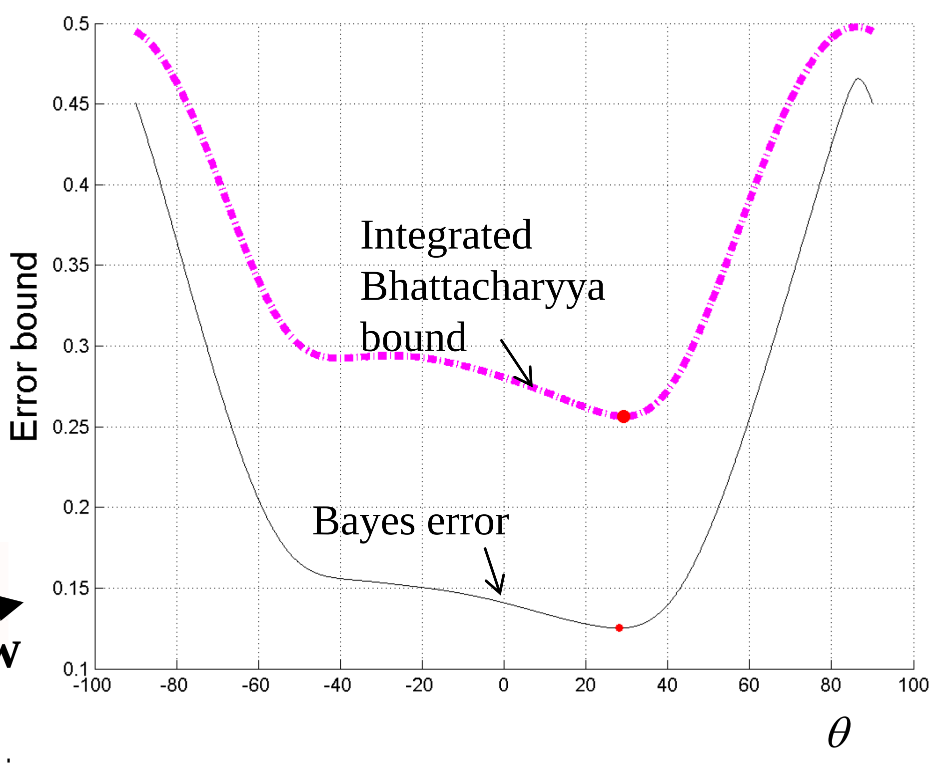
$$J_{Bhatt}(W) = \frac{1}{2}\int p_1^{1/2}(\mathbf{x};W)\, p_2^{1/2}(\mathbf{x};W)\, d\mathbf{x}$$

Characteristic 1:

- Non-convex Function with Global Minimum Near Bayes Optimal.



$$E_{Bayes} \leq J_{Bhatt}$$



Characteristic 2:

- Optimizing Integrated Bhattacharrya Bound.

$$-\ln\left[2J_{Bhatt}(W)\right] = J_1(W) + J_2(W)$$

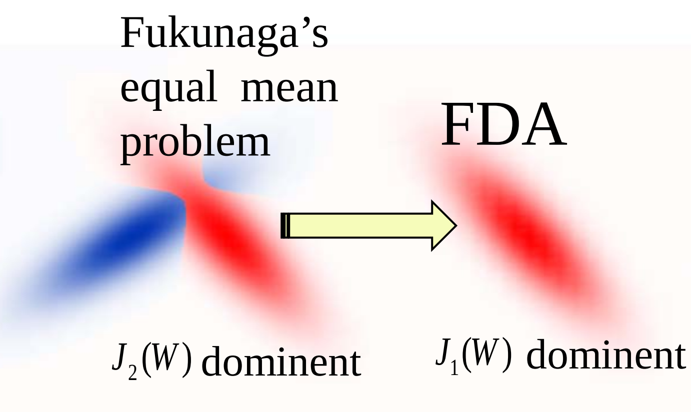
[Fukunaga 1990]

$$J_1(W) = \frac{1}{8} Tr\left[\underbrace{\left(\frac{W^T\Sigma_1 W + W^T\Sigma_2 W}{2}\right)^{-1}}_{\text{Minimizing within class variance}} \underbrace{W^T(\mu_1-\mu_2)(\mu_1-\mu_2)^T W}_{\text{Maximizing mean difference}}\right]$$

μ_i: mean

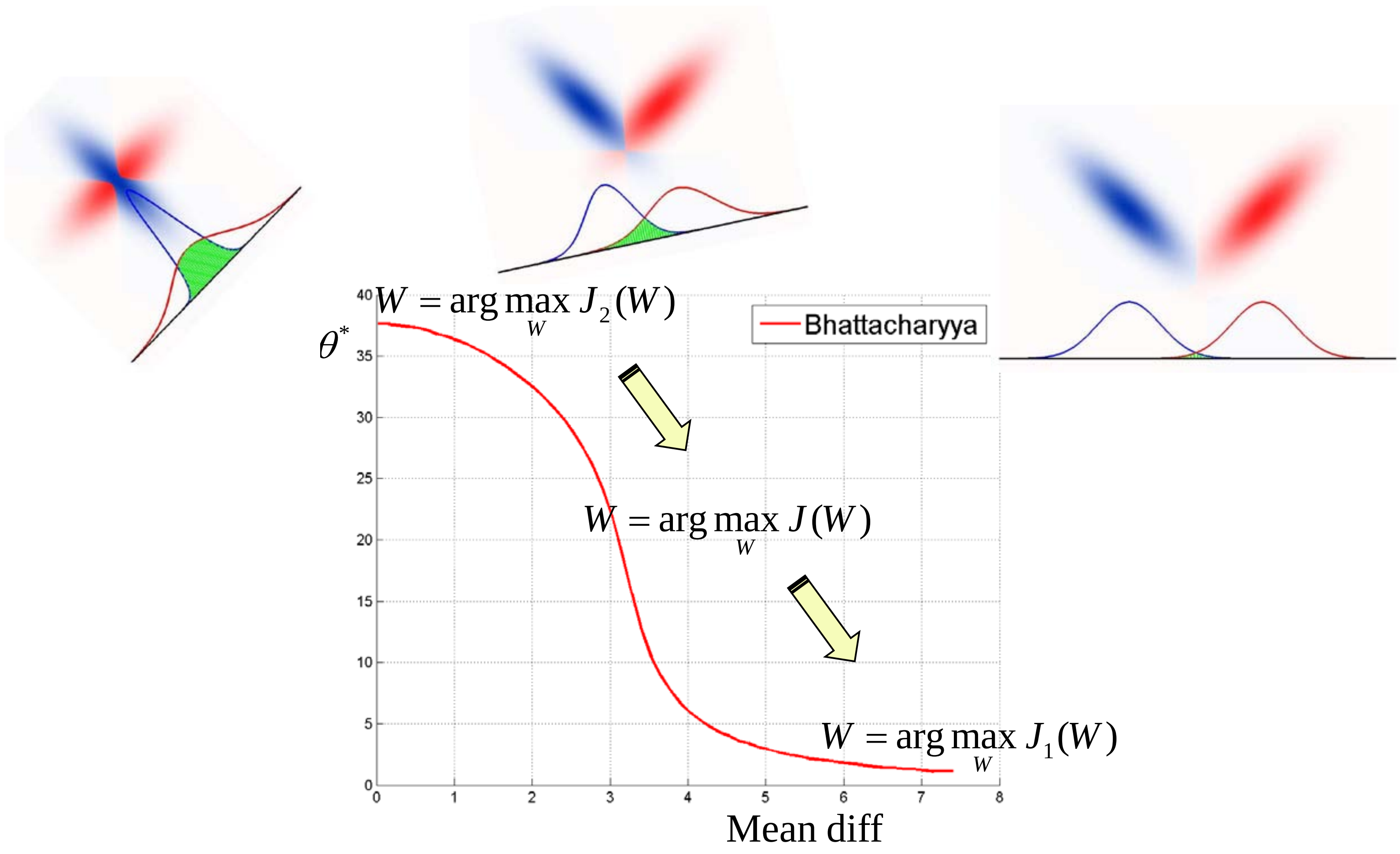
Σ_i: Cov.

$$J_2(W) = \frac{1}{2} \ln \underbrace{\left(\left|\frac{W^T\Sigma_1 W + W^T\Sigma_2 W}{2}\right| / \left|W^T\Sigma_1 W\right|^{1/2} \left|W^T\Sigma_2 W\right|^{1/2} \right)}_{\text{Maximizing covariance difference}}$$



Characteristic 3:

- Transition of the Optimal Direction.



Optimizing Bhattacharrya Bound

- A Good Approximation of the Bayes Optimumal Subspace
- Analytically Integrable for Gaussians

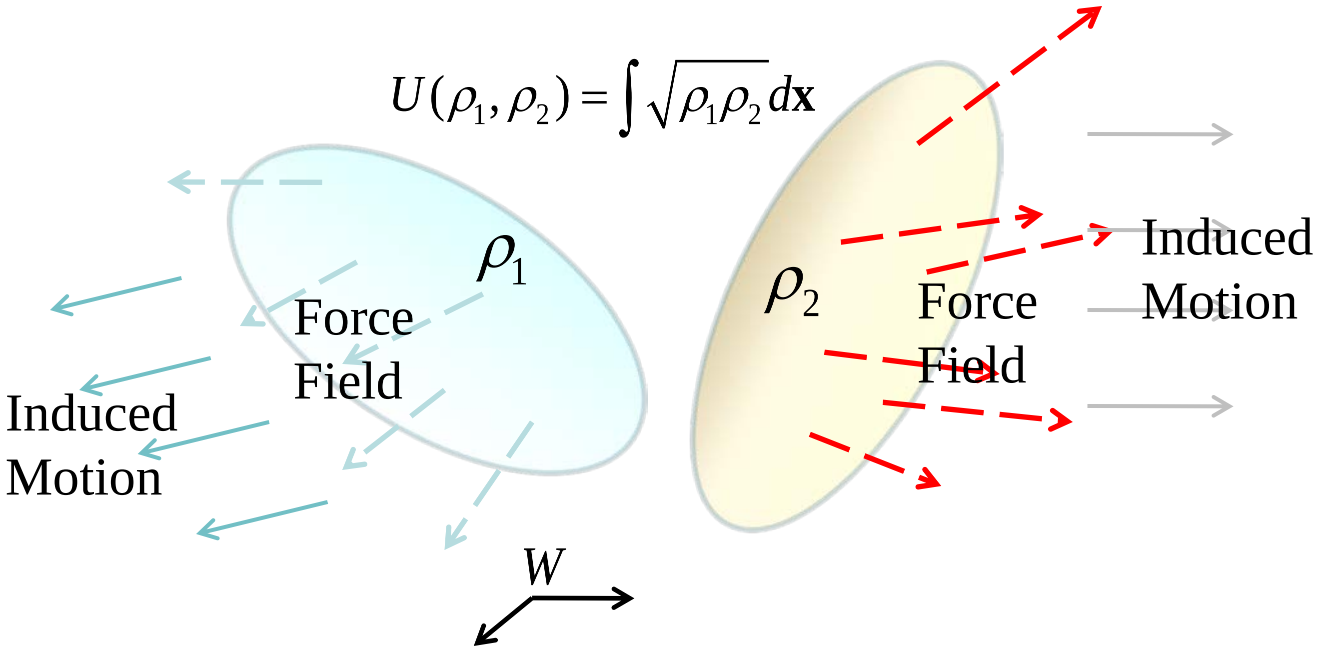
- Non-Convex, Many Local Minima
- ‘*W*’ Matrix Cannot be Optimized by Optimizing Basis Vectors. (Deflation Cannot be Applied.)

How to address these problems?

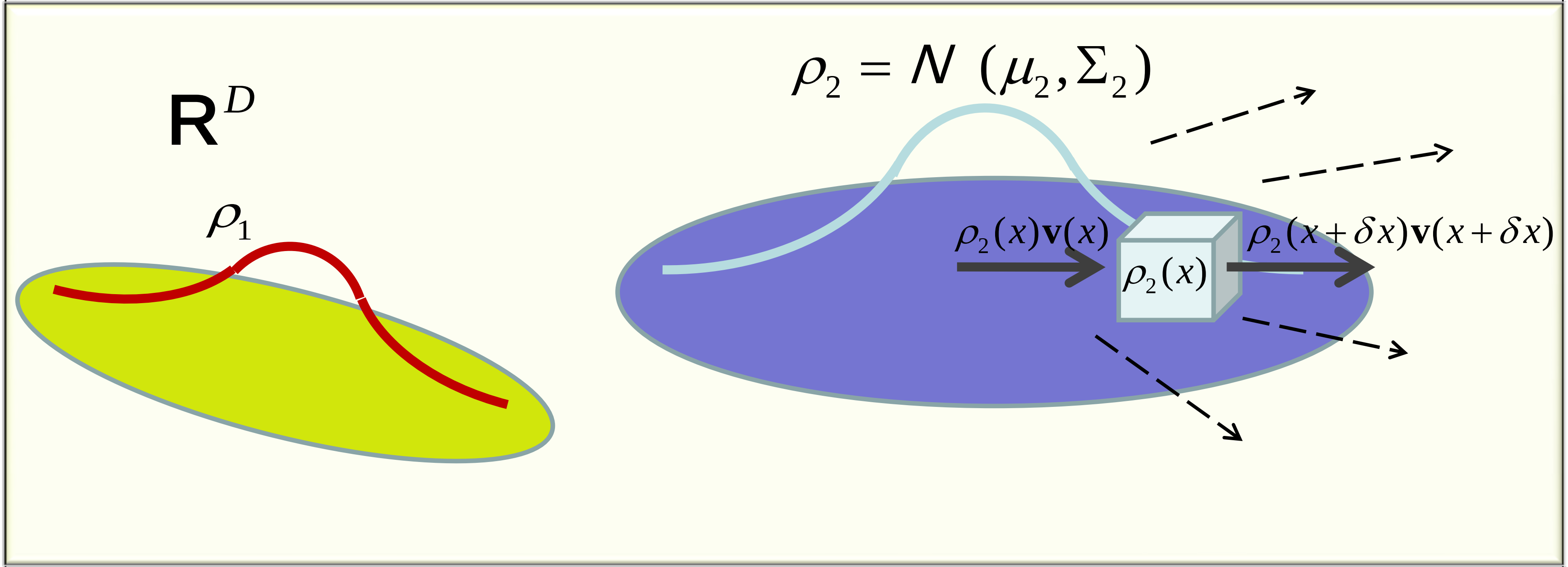
Fluid Dynamics Model

Physics-Based Model for Bhattacharrya Optimization

- Bhattacharyya potential over high dimensional distributions to reduce the overlap
  - Generate the motion
  - Motion & Low dimensional space



Fluid Model

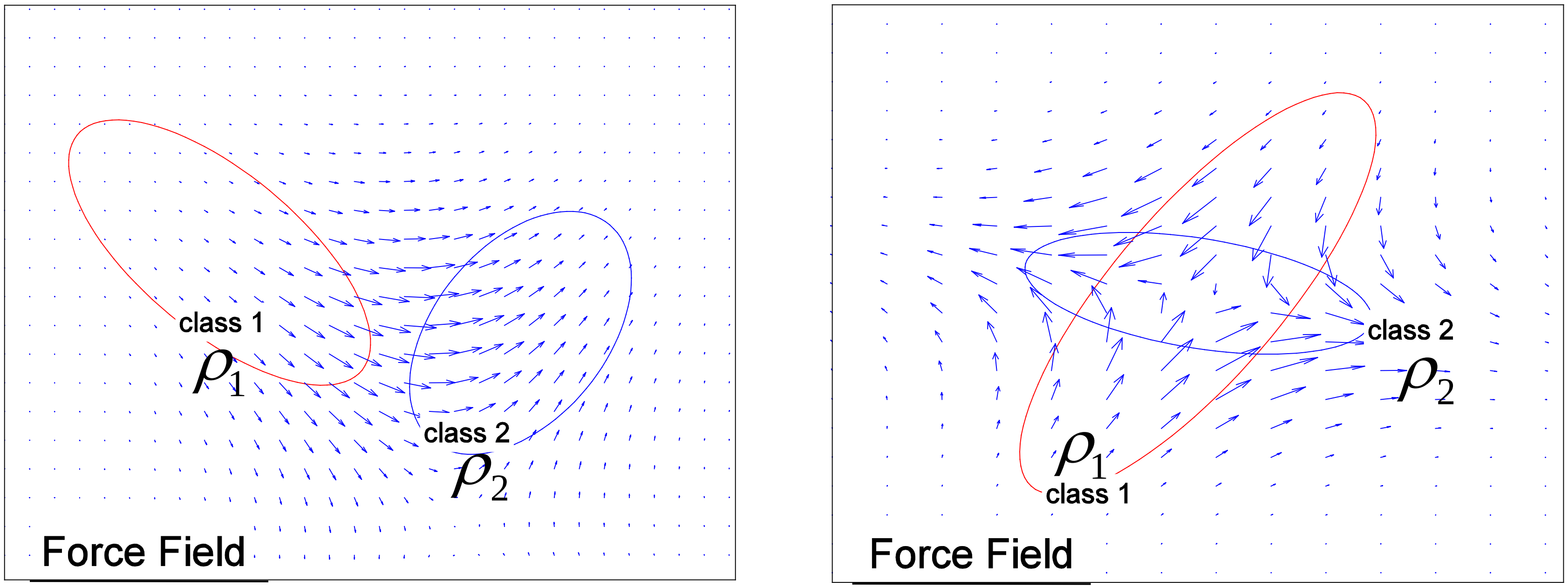


- Equation of continuity

$$\frac{\partial \rho_c}{\partial t} + \nabla \cdot (\rho_c \mathbf{v}) = 0$$

  – Conservation of mass

Force Field for Class Separation



- Potential to reduce interaction

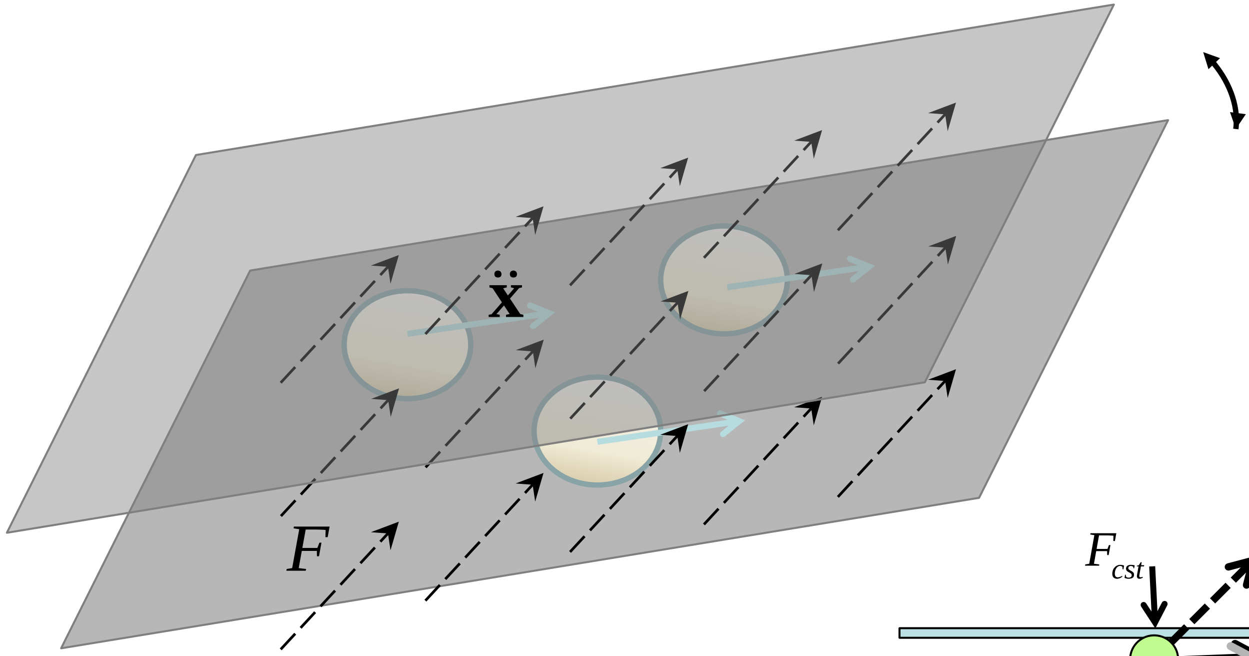
$$U(\rho_1, \rho_2) = \int \sqrt{\rho_1 \rho_2}\, d\mathbf{x}$$ ← Bhatttacharyya bound

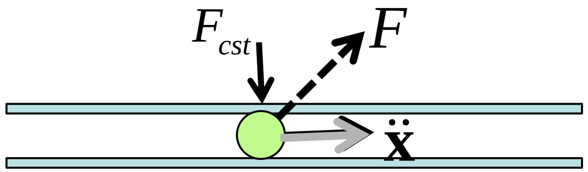
  – Decrease of potential by movement of ρ_2

$$F_{1\to 2}(\mathbf{x}) = -\frac{1}{2}\rho_2 \nabla \sqrt{\frac{\rho_1}{\rho_2}}$$

Constrained Fluid Model

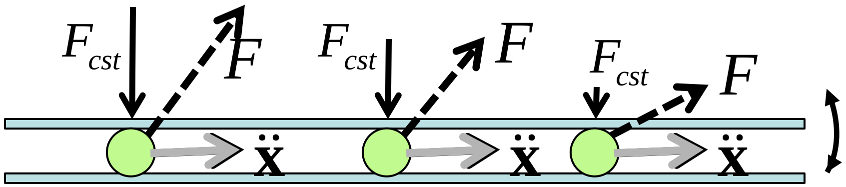
- Acceleration is optimized under force field and given constraint.





F_{cst}: force from constraint

Gauss Principle of Least Constraint



F_{cst}: force from constraint

- Minimize *L*

$$L = \frac{1}{2}\int \rho(\mathbf{x})\left(\ddot{\mathbf{x}} - \frac{F(\mathbf{x})}{\rho(\mathbf{x})}\right)^2 d\mathbf{x}$$

$F_{cst}(\mathbf{x}) / \rho(\mathbf{x})$

  – Minimizing total constraint forces

  – Motion of multiple objects with multiple constraints.

Acceleration Field with Constraint

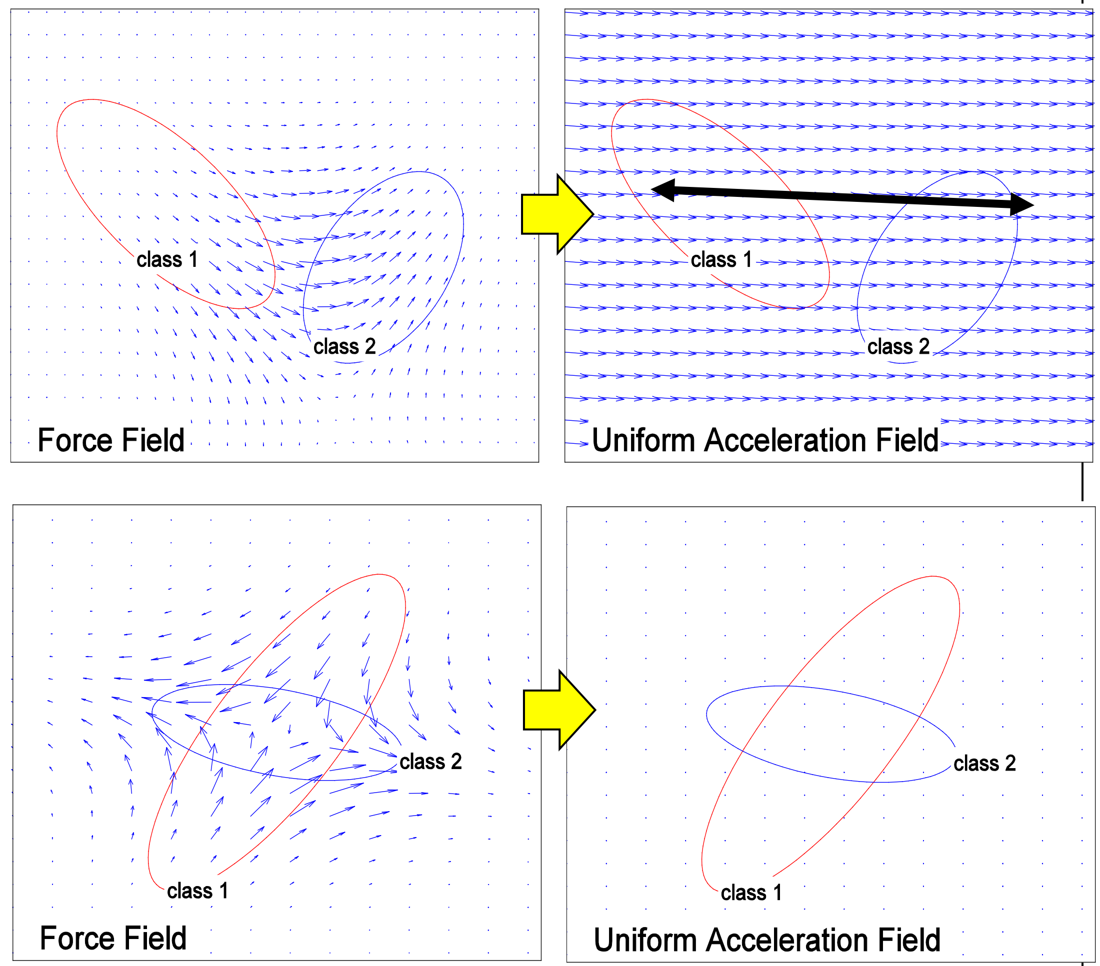
- Constraint on motion

Uniform translational movement

$$\ddot{\mathbf{x}} = \mathbf{w}$$

$$\mathbf{w} = \int F d\mathbf{x}$$
$$= C_2 \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2)$$

The subspace is the FDA solution



Acceleration Field with Constraint

- Constraint on motion

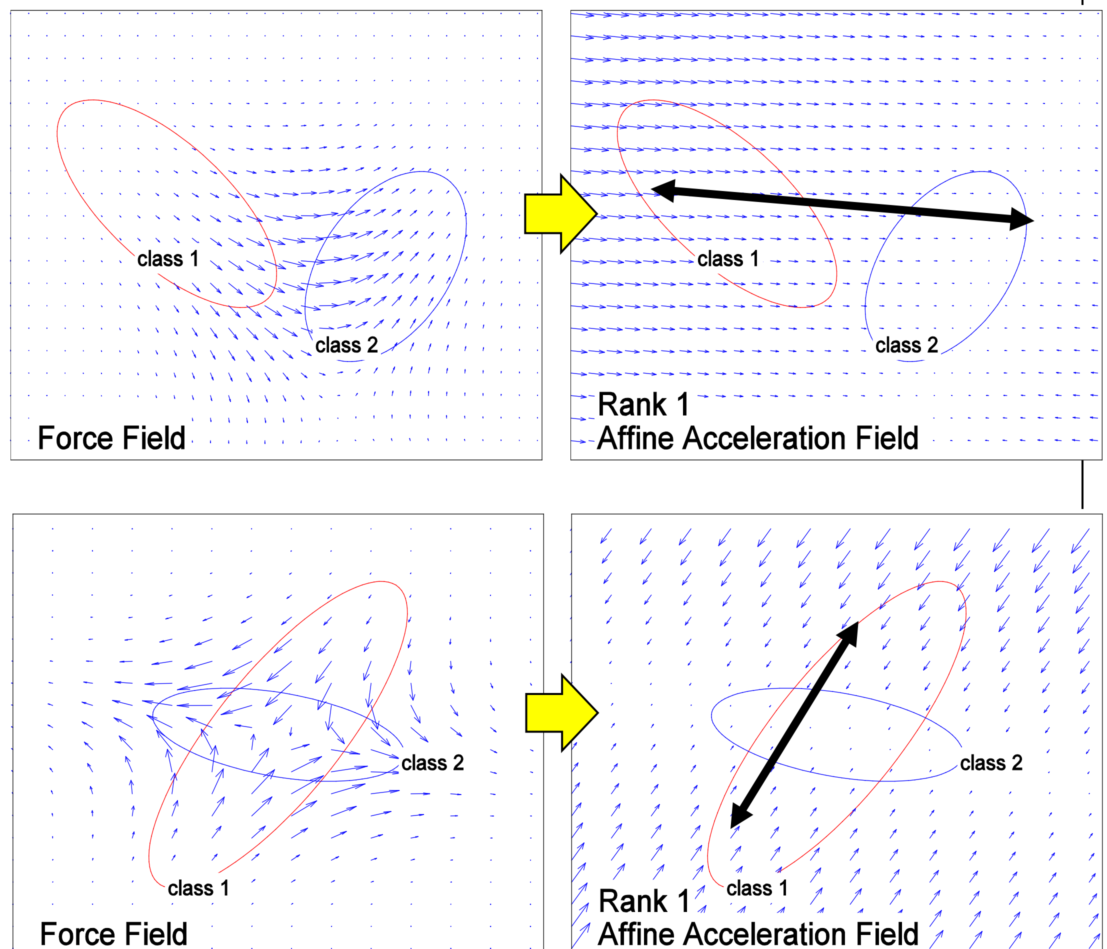
Low rank affine acceleration

$$\ddot{\mathbf{x}} = W U_c^T \mathbf{x}_e \qquad \mathbf{x}_e = \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}$$

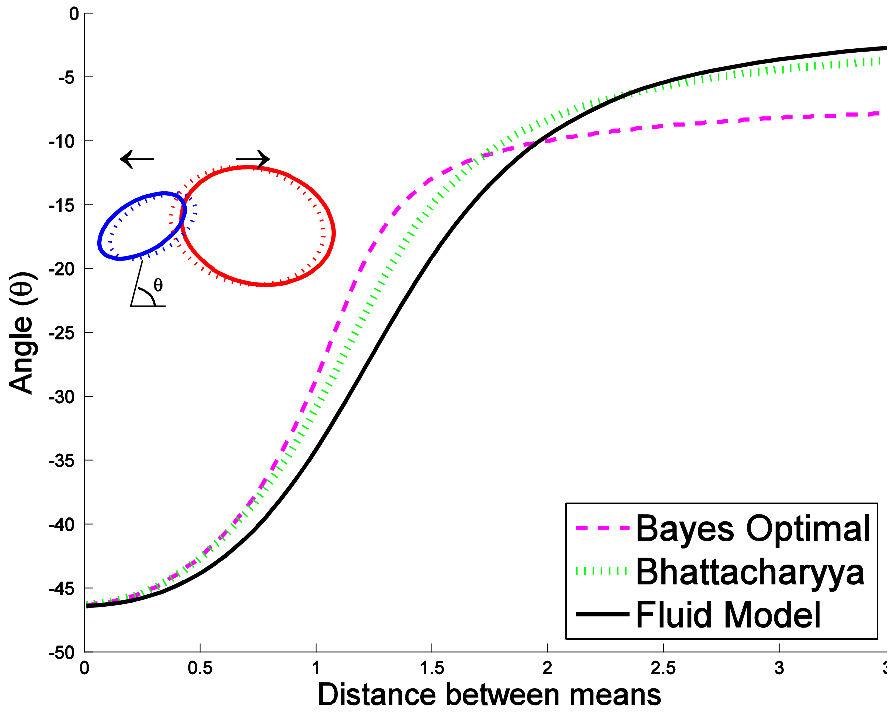
W: e-vectors of

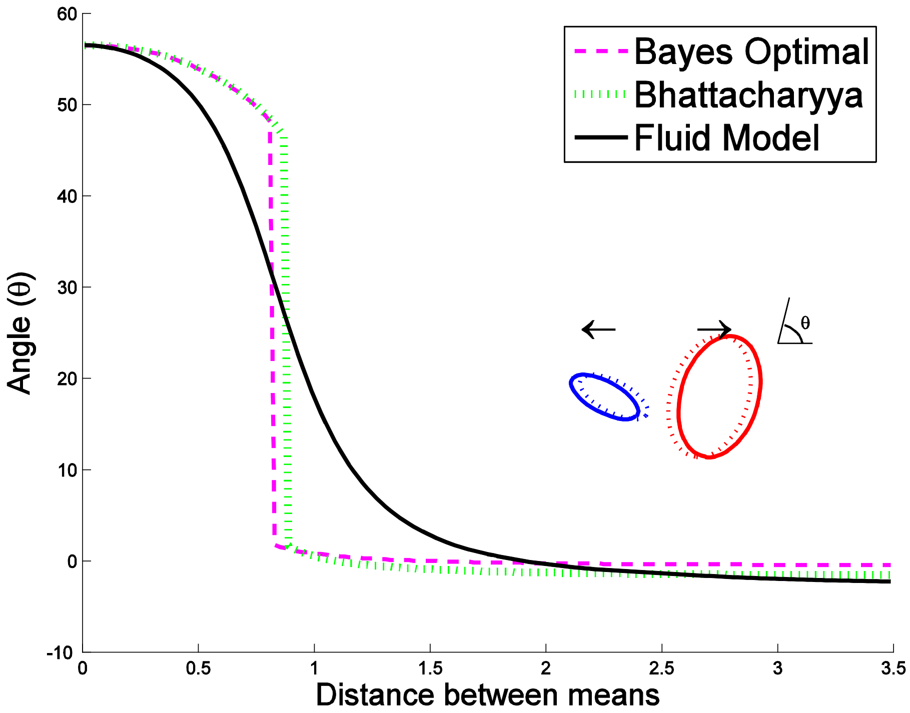
$$\sum_c \left\langle F_c \mathbf{x}_e^T \right\rangle \left\langle \mathbf{x}_e \mathbf{x}_e^T \right\rangle_{\rho_c}^{-1} \left\langle \mathbf{x}_e F_c^T \right\rangle$$

The subspace approximates two analytical Bhattacharyya solution



Optimal Direction for Rank-one Affine Constraint





Real Dataset Performance

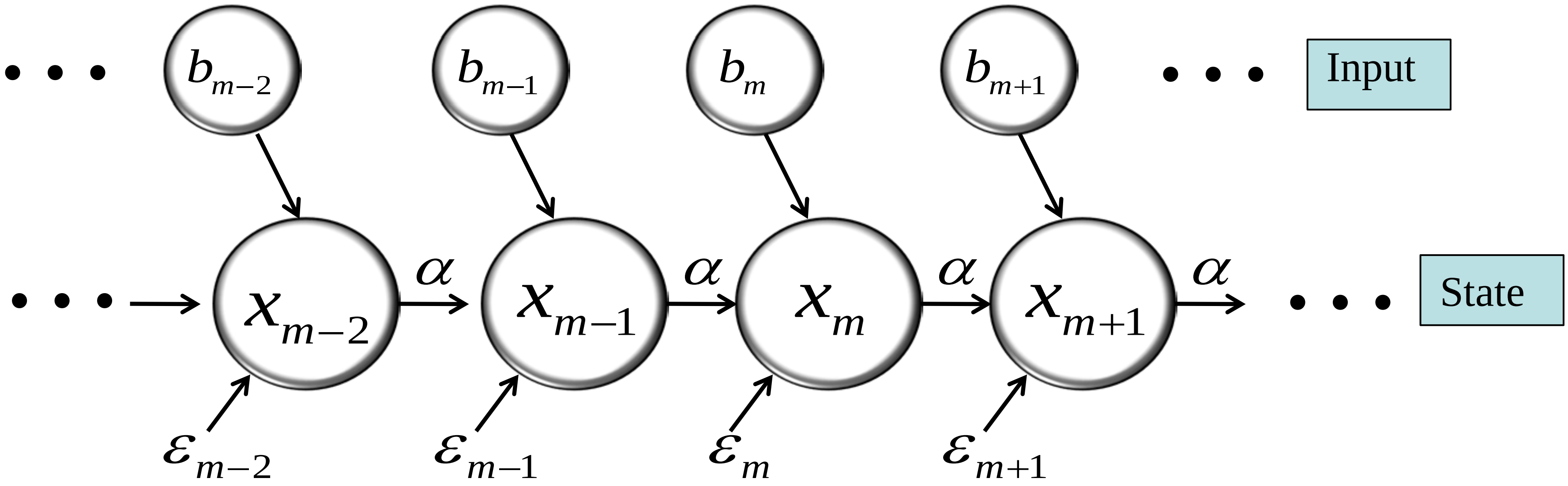
(%)

Dataset	# classes	Fukunaga	FDA	Fluid
SpectF	2	78.50	80.20	**81.70**
Ionosphere	2	86.83	85.96	**87.54**
Parkinsons	2	86.83	82.33	**89.33**
Ozone	2	71.27	**84.54**	84.20
Breast Cancer	2	93.83	97.87	**97.92**
Glass	6	--	52.53	**55.67**
USPS	10	--	90.38	**91.48**

- Average of 100 realizations for UC Irvine datasets
- Average of 20 realizations for USPS dataset
- Number of dimensions of subspace is (number of classes – 1)
- Empirical covariance matrix is regularized with isotropic matrix
- Cross validation is used for each algorithm to choose regularization parameter

Interpretation in Gaussian Processes

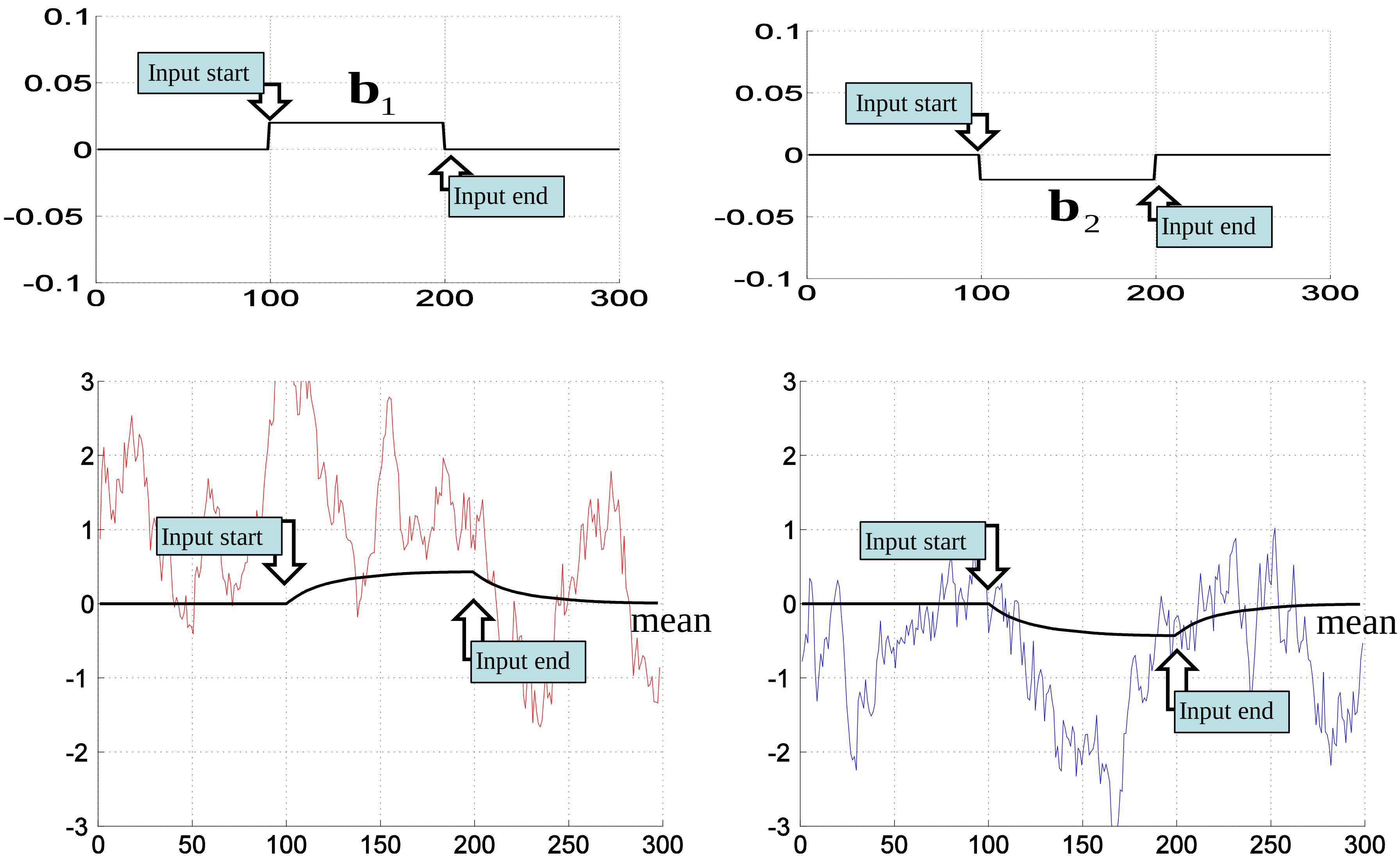
Application to Gaussian Processes



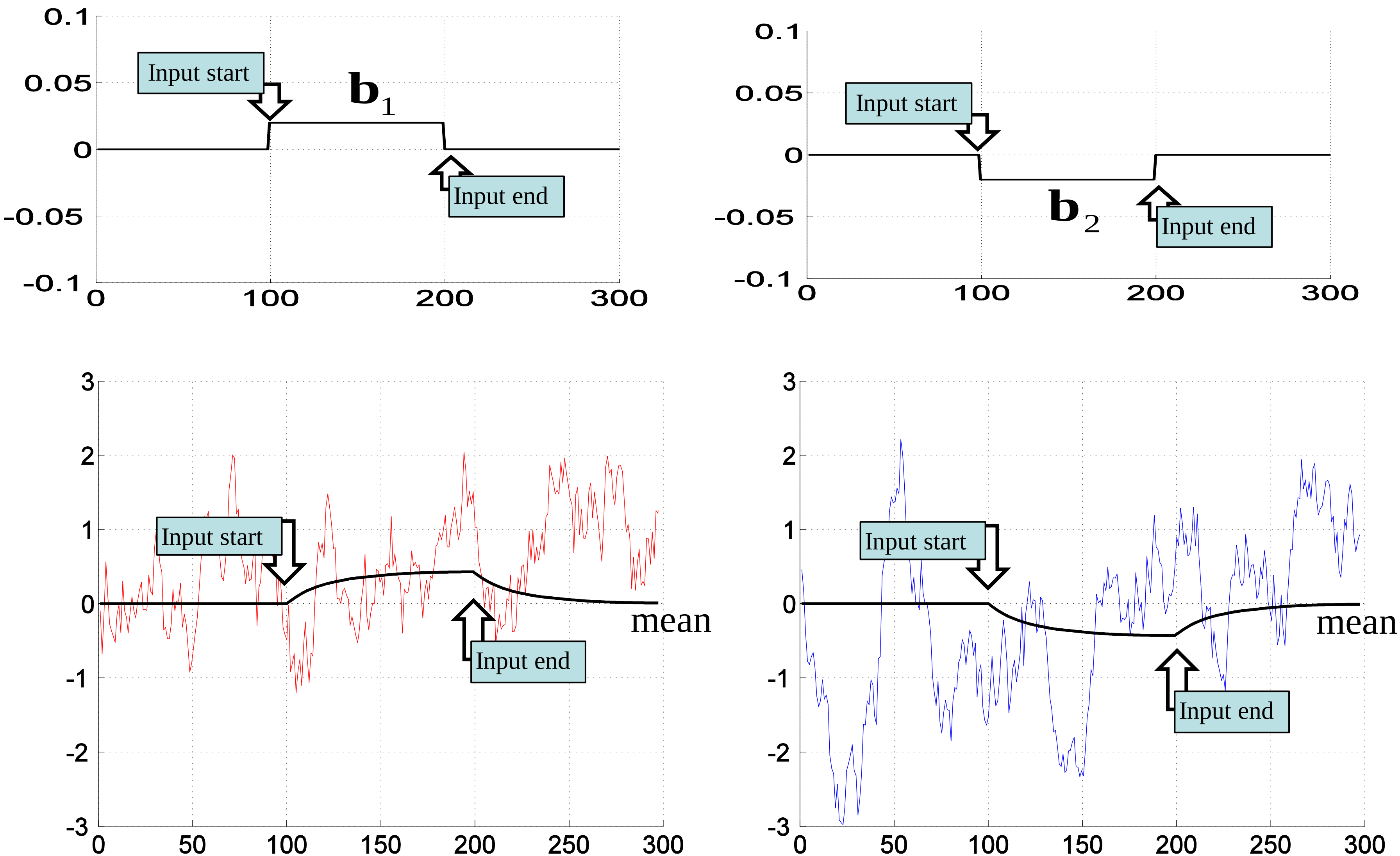
Update rule: $x_m = \alpha x_{m-1} + b_m + \varepsilon_m$

$$0 < \alpha < 1 \qquad \varepsilon_m \sim N\ (0, \sigma^2)$$

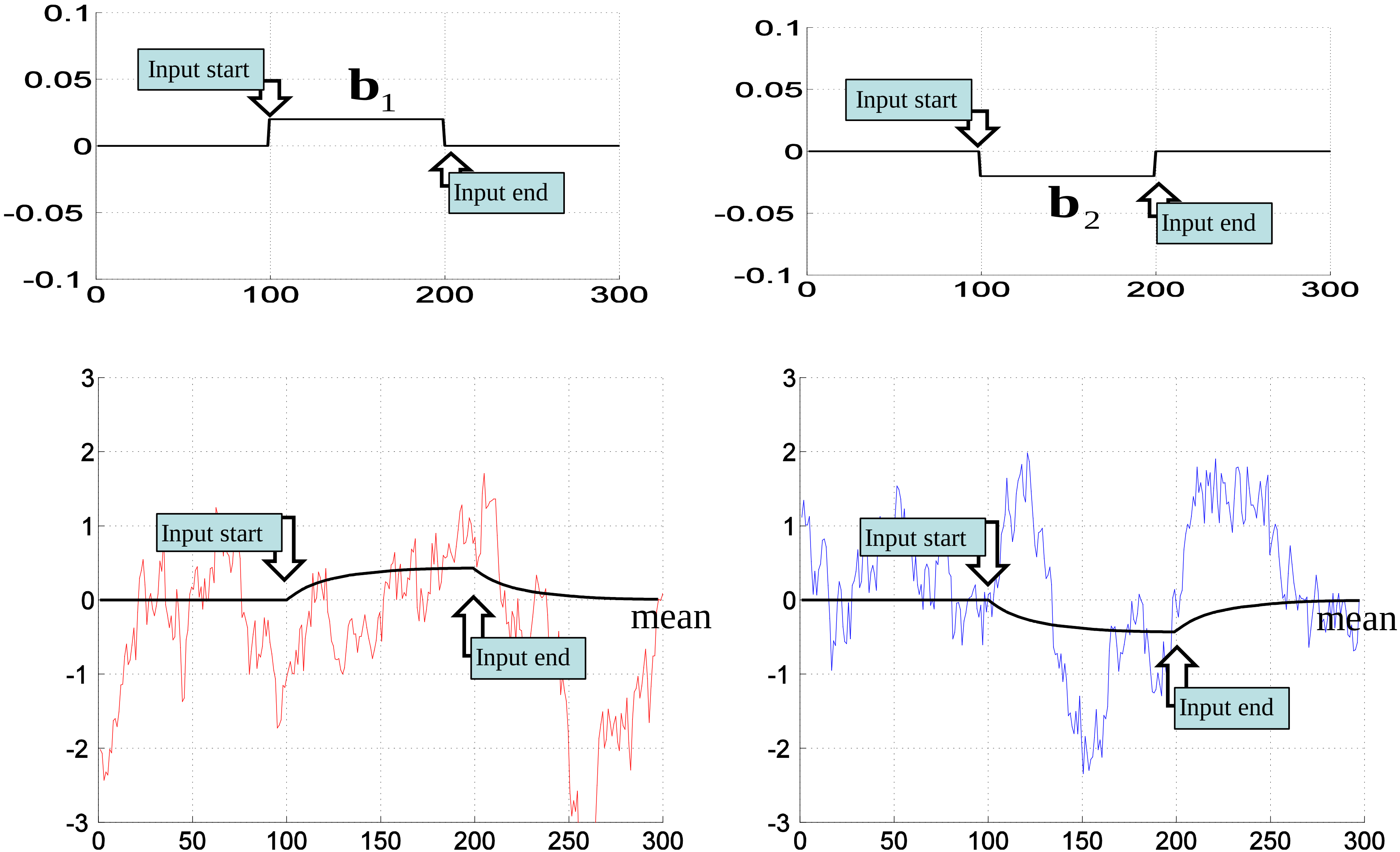
$$GP_1(\mathbf{x} \mid \alpha, \mathbf{b}_1) \qquad GP_2(\mathbf{x} \mid \alpha, \mathbf{b}_2)$$



$GP_1(\mathbf{x} \mid \alpha, \mathbf{b}_1)$ $GP_2(\mathbf{x} \mid \alpha, \mathbf{b}_2)$

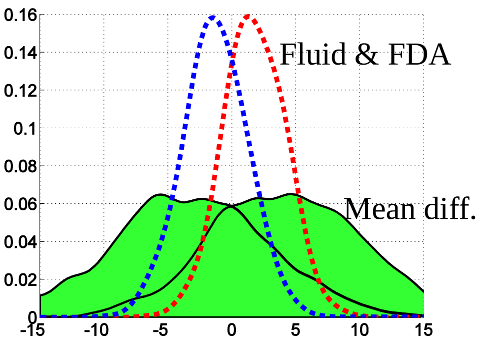


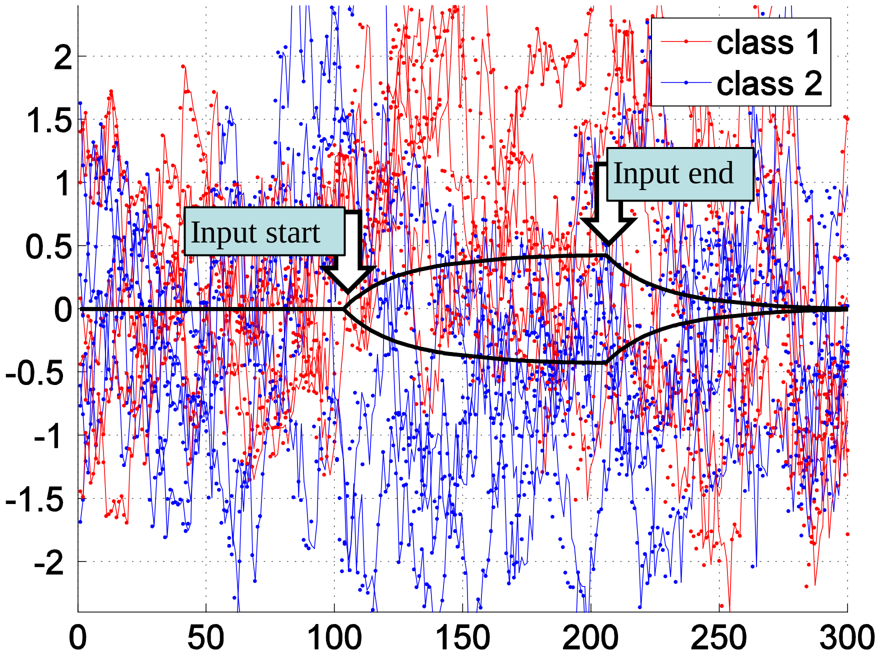
$GP_1(\mathbf{x} \mid \alpha, \mathbf{b}_1)$ $GP_2(\mathbf{x} \mid \alpha, \mathbf{b}_2)$

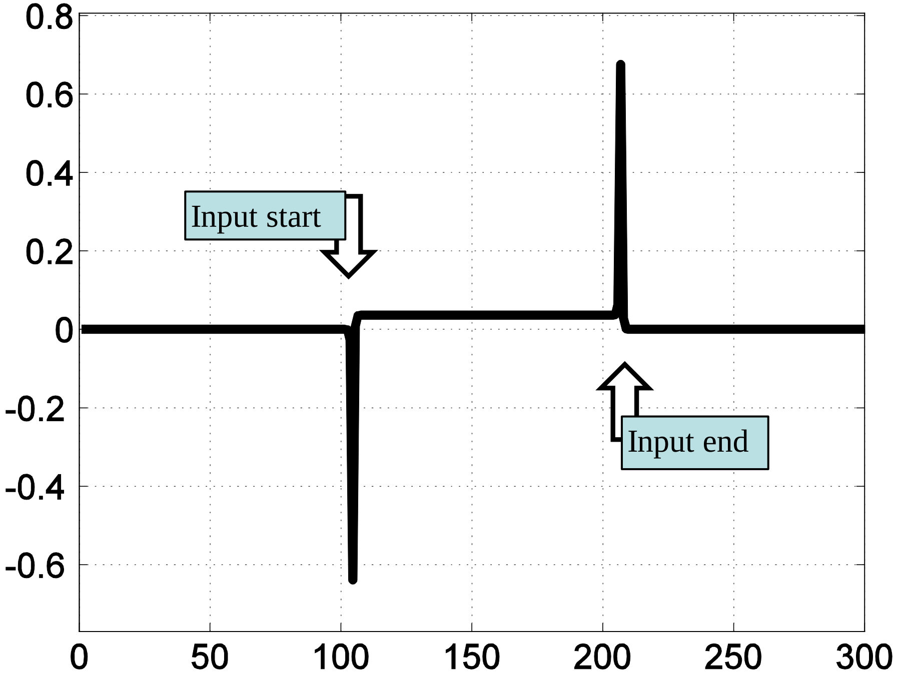


Optimal Filter for GP

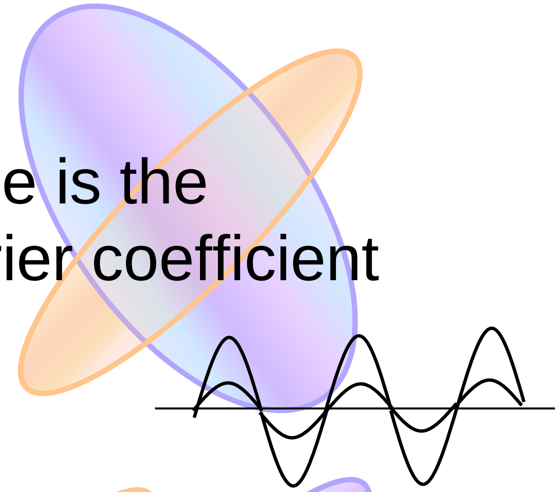
- Same decay rate α
  - Equal cov. function case



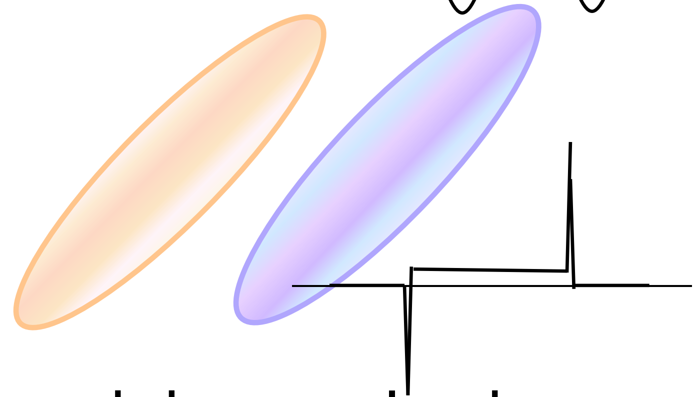




Gaussian Process Interpretation



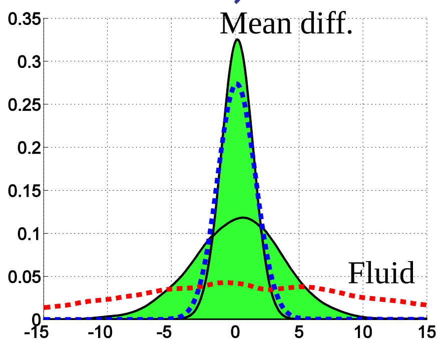
- Equal mean:
  - Filter of maximum variance difference is the ***Fourier mode*** having maximum Fourier coefficient difference.

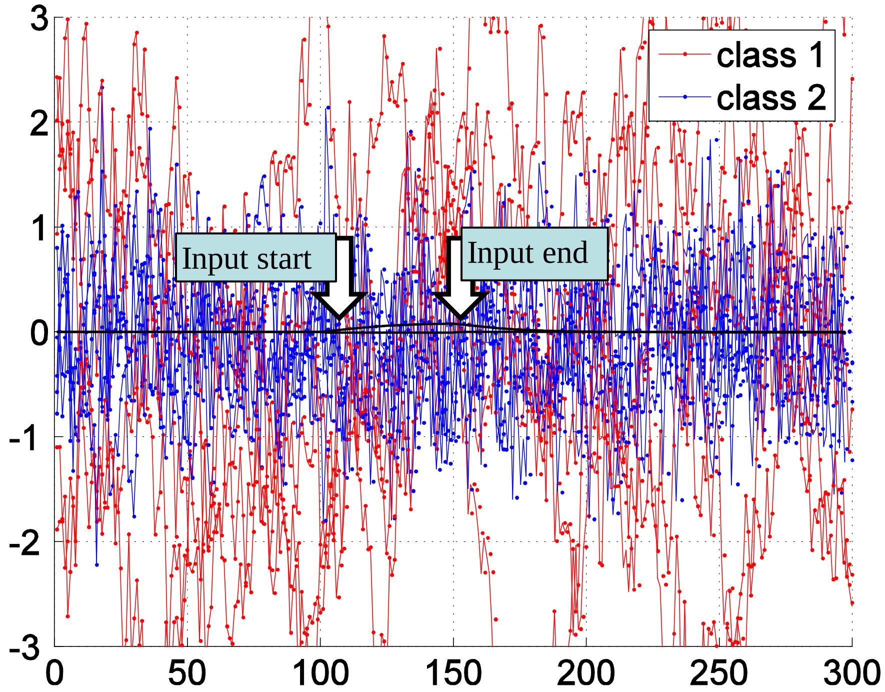


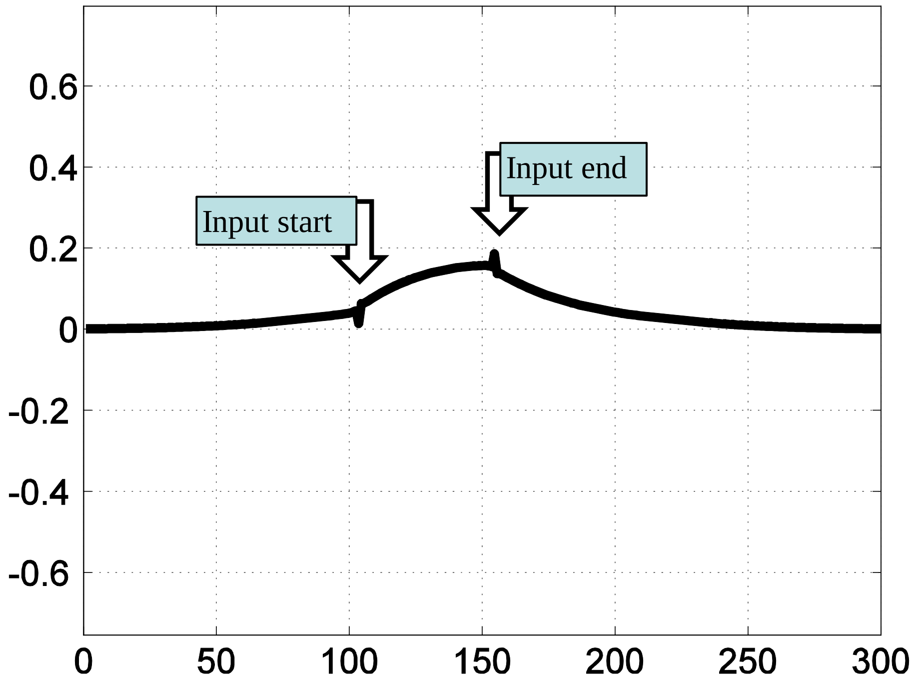
- Equal cov.:
  - Deconvolved input.
- Optimal filter in general:
  - Combination of Fourier modes and deconvolved input (calculated in Fourier domain).

Optimal Filter for GP (Fluid Model)

- Different decay rate α and b.
  - Different mean function
  - Different cov. function







Summary

- Bhattacharyya Analysis
- Physics-based approach (Fluid model) that approximates Bhattacharyya solution.
- Optimal filter approximation for Gaussian Processes.

Thank You

Yung-Kyun Noh

nohyung@seas.upenn.edu