ICASSP2006, Toulouse, France May 14-19, 2006

Obtaining the Best Linear Unbiased Estimator of Noisy Signals by Non-Gaussian Component Analysis

M. Sugiyama[1], M. Kawanabe[2], G. Blanchard[2], V. Spokoiny[3] and K.-R. Müller[2,4]

1 Tokyo Institute of Technology, Japan
2 Fraunhofer FIRST.IDA, Germany
3 Weierstrass Institute & Humboldt University, Germany
4 University of Potsdam, Germany

Signal Denoising

- Signals we observe in practice are often noisy and redundant.

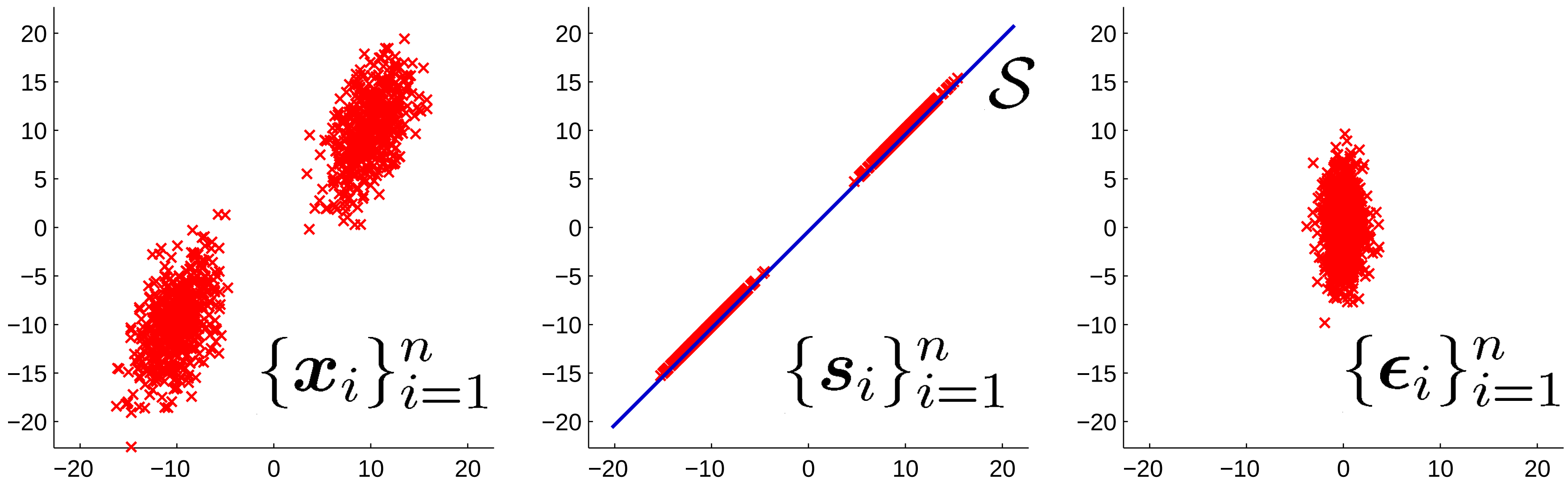
Observed signal $\boldsymbol{x}$	=	Low-dim. signal $\boldsymbol{s}$	+	Full-dim. noise $\boldsymbol{\epsilon}$
$\boldsymbol{x} \in \mathbb{R}^d$		$\boldsymbol{s} \in \mathcal{S} \subset \mathbb{R}^d$		$\boldsymbol{\epsilon} \in \mathbb{R}^d$

$\mathcal{S}$: signal subspace

- We want to remove noise $\boldsymbol{\epsilon}$ by cleverly making use of signal redundancy

Setting

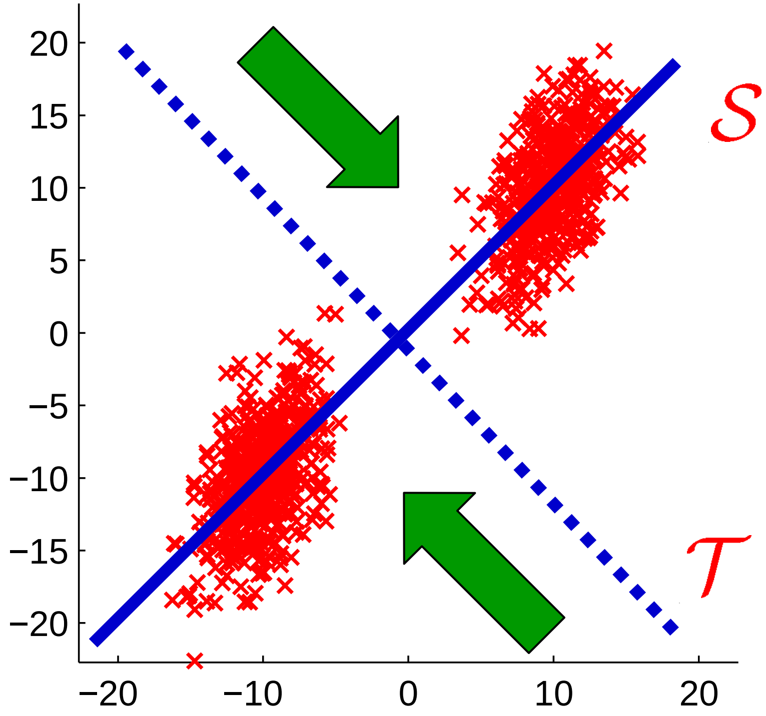
- True signal is non-Gaussian: $\boldsymbol{s} \sim p(\boldsymbol{s})$
- Noise is centered Gaussian: $\boldsymbol{\epsilon} \sim \phi(\boldsymbol{\epsilon})$
- $\boldsymbol{s}$ and $\boldsymbol{\epsilon}$ are statistically independent.
- We observe i.i.d. noisy samples: $\{\boldsymbol{x}_i\}_{i=1}^n$



Goal: obtain good estimates $\{\widehat{\boldsymbol{s}}_i\}_{i=1}^n$ of $\{\boldsymbol{s}_i\}_{i=1}^n$

Typical Denoising Strategy

Project $\{\boldsymbol{x}_i\}_{i=1}^{n}$ onto $\mathcal{S}$



$\widehat{\boldsymbol{s}}_i = \boldsymbol{P}\boldsymbol{x}_i$

$\boldsymbol{P}$:orthogonal projection

- Projection does not have to be orthogonal.
- We want to choose "along"-subspace $\mathcal{T}$ so that noise is maximally reduced.

Best Linear Unbiased Estimator (BLUE)

- Linear estimator: $\widehat{\boldsymbol{s}}_i = \boldsymbol{H}\boldsymbol{x}_i$
- Unbiased estimator: $\mathbb{E}_{\boldsymbol{\epsilon}}[\widehat{\boldsymbol{s}}_i] = \boldsymbol{s}_i$

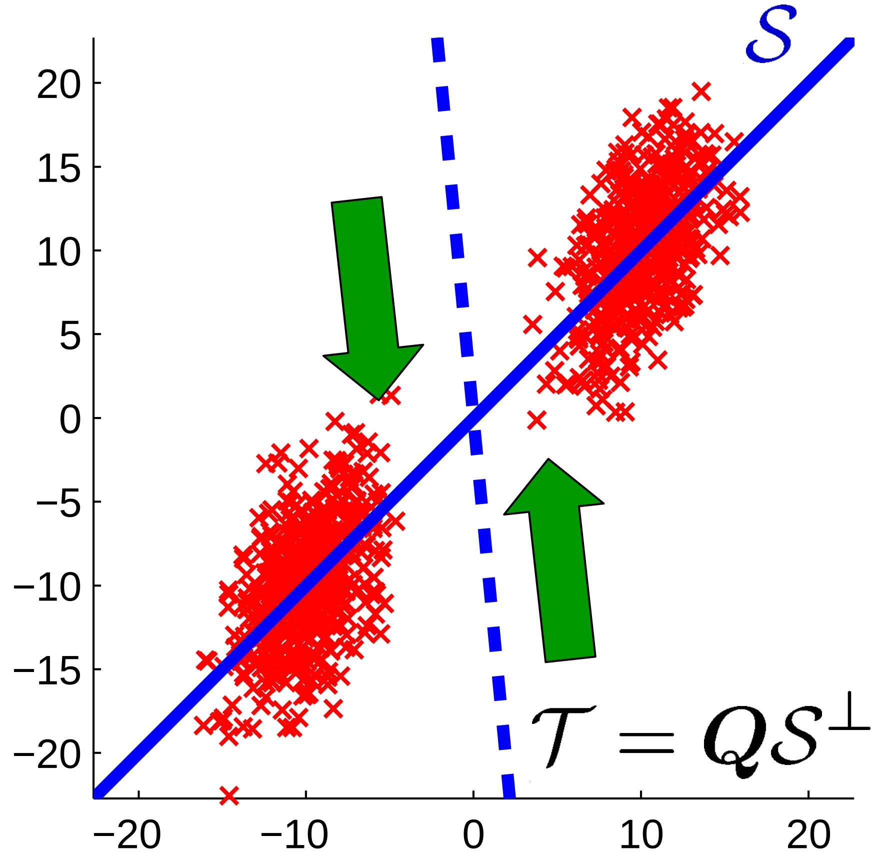
  $\mathbb{E}_{\boldsymbol{\epsilon}}$: Expectation over noise
- BLUE: Minimum variance estimator among all linear unbiased estimators

$$\mathbb{E}_{\boldsymbol{\epsilon}}(\widehat{\boldsymbol{s}}_i - \mathbb{E}_{\boldsymbol{\epsilon}}[\widehat{\boldsymbol{s}}_i])^2 \le \mathbb{E}_{\boldsymbol{\epsilon}}(\widetilde{\boldsymbol{s}}_i - \mathbb{E}_{\boldsymbol{\epsilon}}[\widetilde{\boldsymbol{s}}_i])^2$$

for any linear unbiased estimator $\widetilde{\boldsymbol{s}}_i$

Geometric View of BLUE

Project $\{\boldsymbol{x}_i\}_{i=1}^{n}$ onto $\mathcal{S}$ along $\mathcal{T} = \boldsymbol{Q}\mathcal{S}^{\perp}$



$\boldsymbol{Q} = \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\top}]$

:Noise covariance matrix

Drawbacks of BLUE

- BLUE can be computed by

$$\widehat{\boldsymbol{s}}_i = \boldsymbol{H}\boldsymbol{x}_i \qquad \boldsymbol{H} = (\boldsymbol{P}\boldsymbol{Q}^{-1}\boldsymbol{P})^{\dagger}\boldsymbol{Q}^{-1}$$

$\dagger$: Moore-Penrose generalized inverse

- Thus we need
  - (A) Noise covariance matrix $\boldsymbol{Q}$
  - (B) Projection matrix $\boldsymbol{P}$ (i.e., need to know $\mathcal{S}$)

- However, $\boldsymbol{Q}$ and $\boldsymbol{P}$ are often unknown.

$\Rightarrow$ **Need to approximate BLUE**

To Cope with (A)

Lemma: $\boldsymbol{Q}$ can be replaced with $\boldsymbol{\Sigma} = \mathbb{E}_{\boldsymbol{x}}[\boldsymbol{x}\boldsymbol{x}^\top]$

$$\boldsymbol{H} = (\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})^\dagger \boldsymbol{\Sigma}^{-1}$$

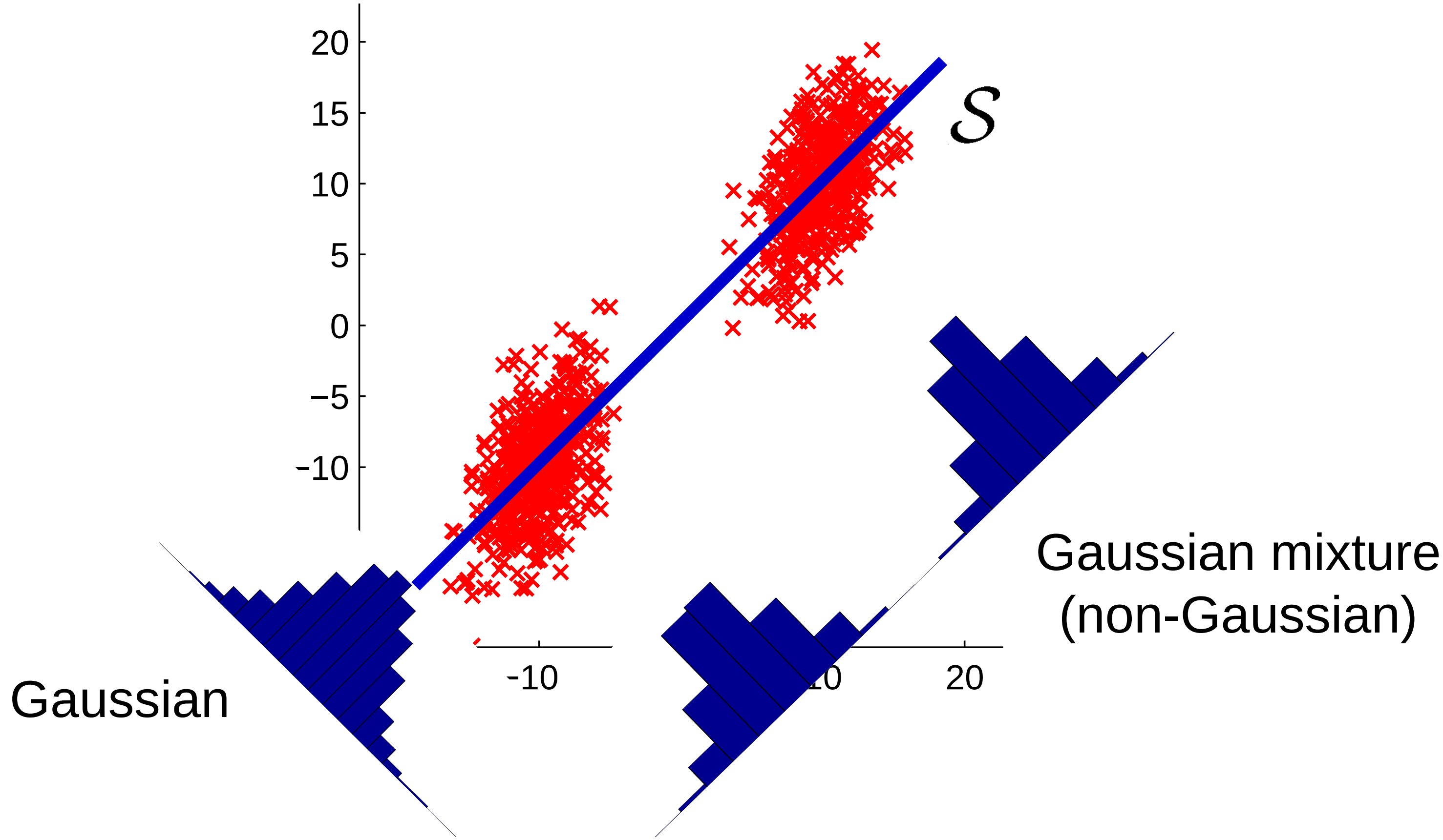
- Intuition:
  - $\boldsymbol{H}$ only affects components in $\mathcal{T}$
  - "$\mathcal{T}$"-part of $\boldsymbol{Q}$ agrees with $\boldsymbol{\Sigma}$ since $\boldsymbol{s} \in \mathcal{S}$
- Utility: Estimating $\boldsymbol{Q}$ is not straightforward, but a consistent estimator of $\boldsymbol{\Sigma}$ can be constructed as

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top$$

To Cope with (B)

Lemma: Non-Gaussian directions in signals is the signal subspace

(Blanchard et al., 2006)



Projection Pursuit

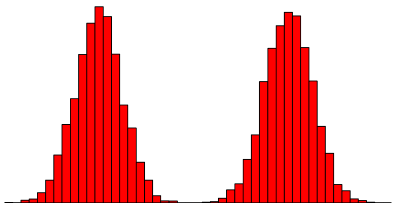
(Friedman & Tukey, 1975)

- Iteratively finding non-Gaussian directions:

$$\widehat{\boldsymbol{\beta}} = \mathop{\mathrm{argmax}}_{\|\boldsymbol{\beta}\|=1} \left| \mathbb{E}_{\boldsymbol{x}}[G(\boldsymbol{\beta}^\top \boldsymbol{x})] - \mathbb{E}_{\nu}[G(\nu)] \right|$$

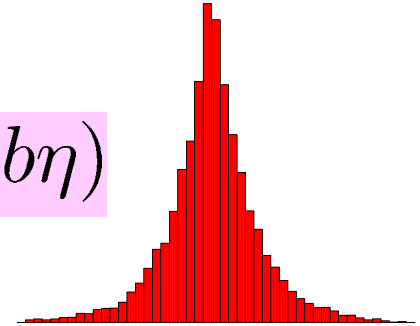
G : projection index $\qquad \nu \sim \mathcal{N}(0,1)$

- Kurtosis: $G_1(\eta) = \eta^4$
  (good for finding sub-Gaussians)



Gaussian mixture

- Robust index: $G_2(\eta) = \frac{1}{b} \log\cosh(b\eta)$
  (good for finding super-Gaussians)

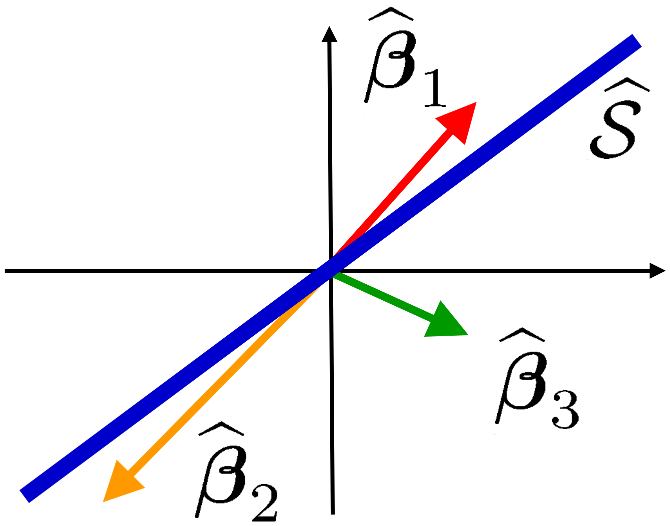


Laplacian

Multi-Index Projection Pursuit

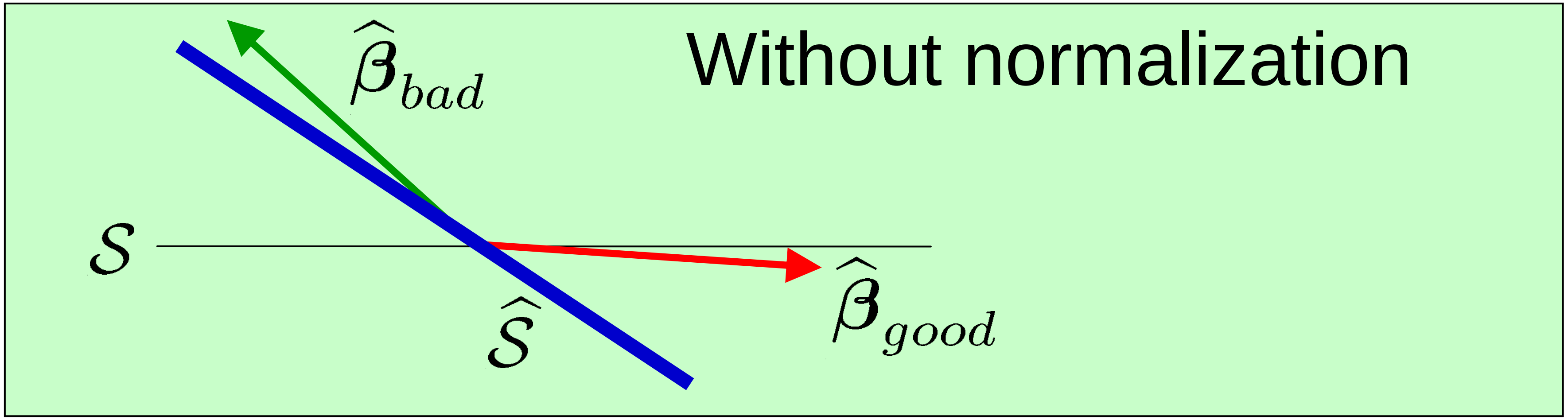
(Blanchard et al., 2006)

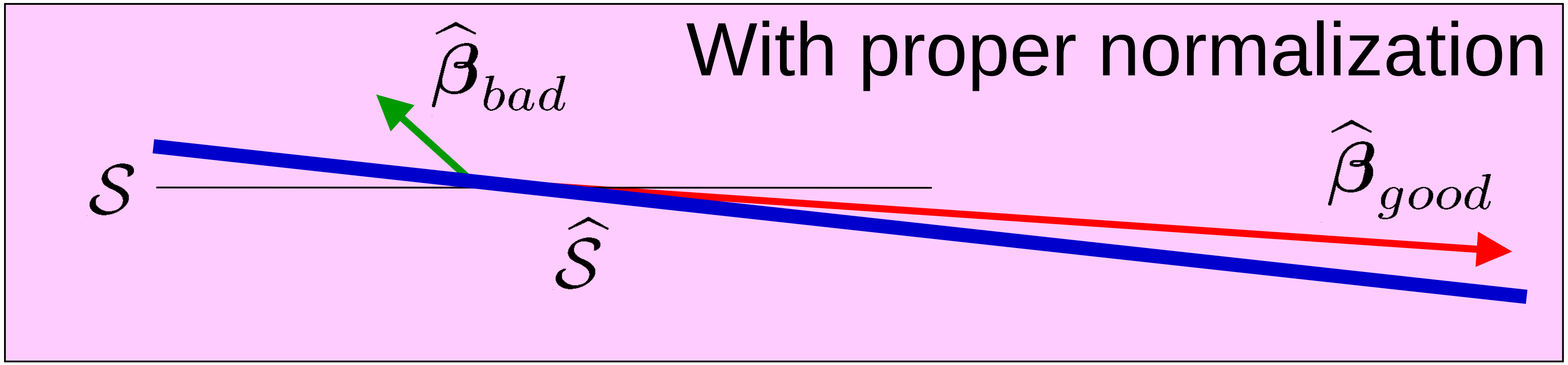
- Performance of PP depends on the choice of projection index.
- If samples contain both super- and sub-Gaussians, no single best index exists.
- MIPP combines results $\{\widehat{\boldsymbol{\beta}}_i\}$ of PP with many different indices by PCA.



Normalization of PP Results

- PCA result becomes reliable if
  - "good" $\widehat{\boldsymbol{\beta}}_i$ has larger norm
  - "bad" $\widehat{\boldsymbol{\beta}}_i$ has smaller norm

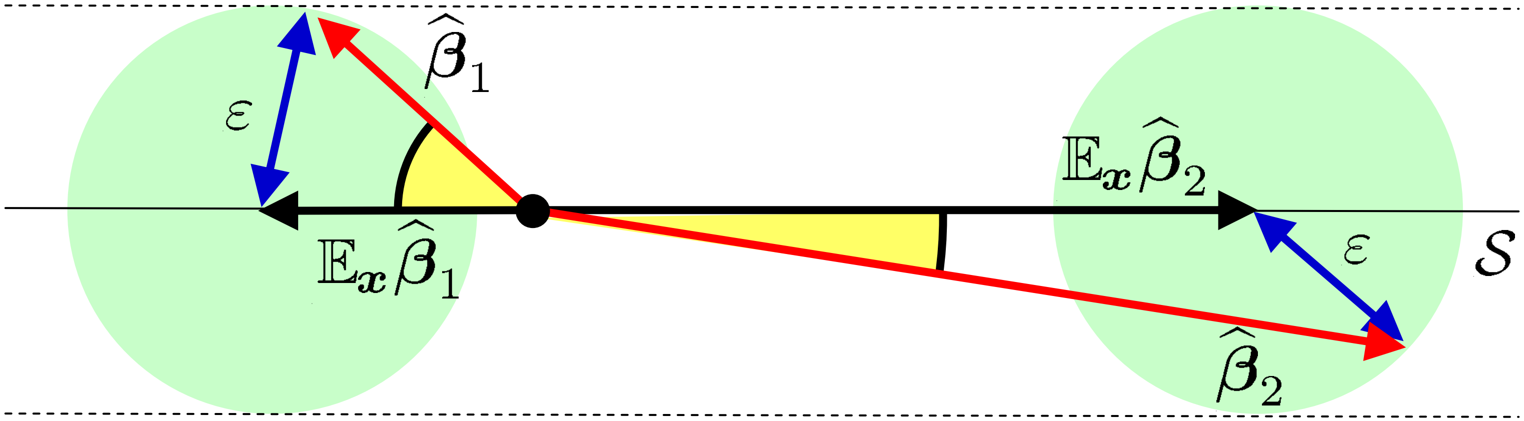




Normalization (cont.)

- Such normalization is achieved by equalizing error orders of $\{\widehat{\boldsymbol{\beta}}_i\}$:

$$\varepsilon_i^2 = \mathbb{E}_{\boldsymbol{x}} \|\widehat{\boldsymbol{\beta}}_i - \mathbb{E}_{\boldsymbol{x}} \widehat{\boldsymbol{\beta}}_i\|^2$$



- In practice, ε_i^2 is approximated by a consistent estimator:

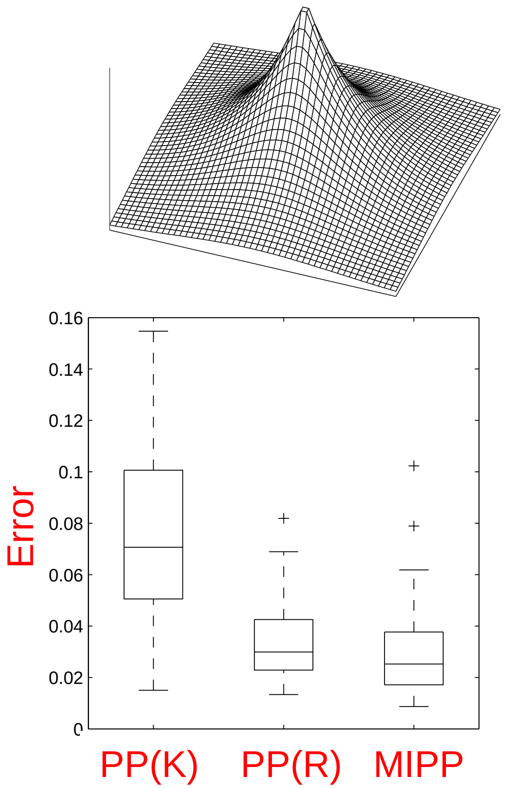
$$\widehat{\varepsilon}_i^2 = \frac{1}{n^2}\sum_{j=1}^{n} \|g_i(\boldsymbol{x}_j)\|^2 - \frac{1}{n}\|\frac{1}{n}\sum_{j=1}^{n} g_i(\boldsymbol{x}_j)\|^2$$

$$g_i(\boldsymbol{x}) = \boldsymbol{x} G_i'(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}) - \boldsymbol{\beta} G_i''(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x})$$

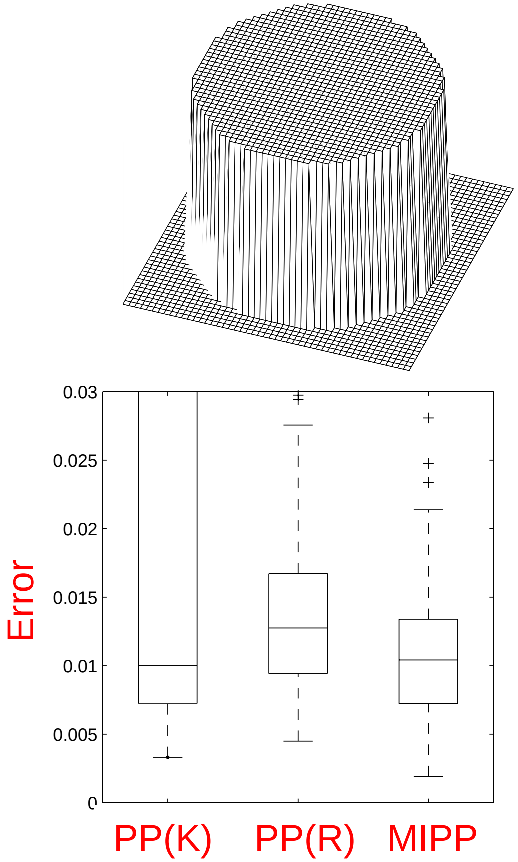
Simulations

- 2-dim. signal & 8-dim. noise

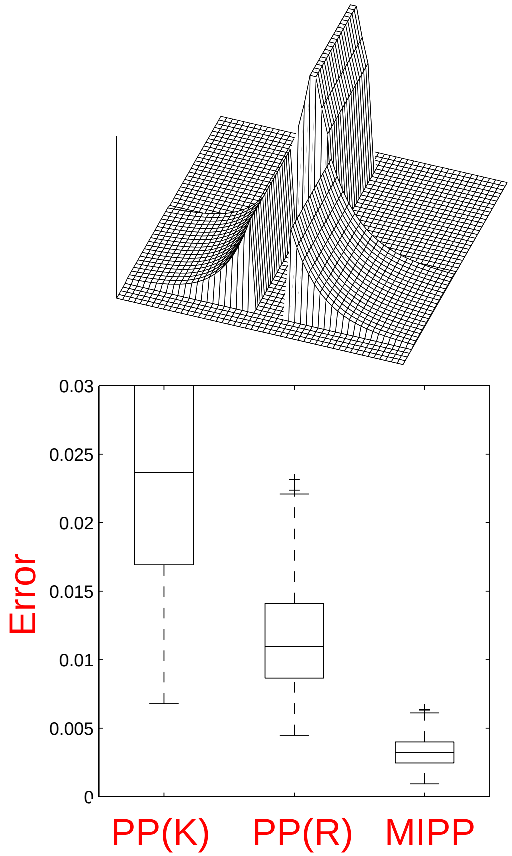
Super-Gaussian



Sub-Gaussian



Super&sub-Gaussian



$$\text{Error} = \|(\boldsymbol{I}_d - \boldsymbol{P})\widehat{\boldsymbol{P}}\|_{Fro}^2$$