Wiener System Identification by Weighted Principal Component Analysis

Q. Zhang **V. Laurain**

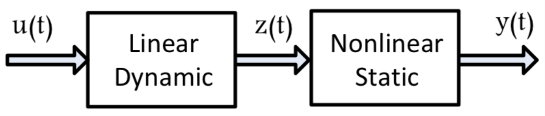
14 juin 2013
vincent.laurain@univ-lorraine.fr

Outline

Identification of Wiener systems

What is a Wiener system ?

- Dynamical block nonlinear systems
- A dynamical linear model followed by a nonlinear function



Two philosophies

- The system is truly, physically, Wiener
  - we know something about the nonlinearities
  - parity, monotony, parametrization
- Wiener models are a good approximation of true systems
  - Nonlinearities represent the unknow/too complex phenomena
  - We do not know anything about the nonlinearities

Truly Wiener system : the race to optimality

Parametric methods

- Heavy optimization algorithms
  - ML, PEM
  - Gradient descent, particule filters
  - Ljung
- How to initialize ?
  - Either not discussed
  - Or use simple LS using FIR

Nonparametric methods

- Non-parametric nonlinearity modelling
  - SVM, Volterra series,...
  - paremeter number increases exponentially with the linear parameter number
  - In Wiener models : often applied on less than 4 parameter FIR examples
  - Grieblicky, Pawlack, Bernstein

Wiener model as system approximation

What is then optimality ?

- Should we really try to identify the nonlinearity ?
- Why not trying to get the best linear approximation possible ?
  - Ideal for control
  - If not precise enough, it can be used for initialisation

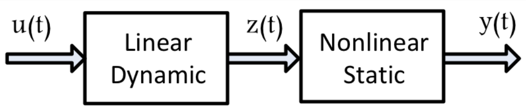
Possible solutions

- Pintelon, Schoukens
  - Either using Frequency identification $\rightarrow$ BLA
  - Or using a FIR approximation $\rightarrow$ LS
- Bai
  - Use FIR approximation
  - Requires only monotony information
- **All these methods are computationally very cost effective**

Wiener Identification seems to go through FIR approximation

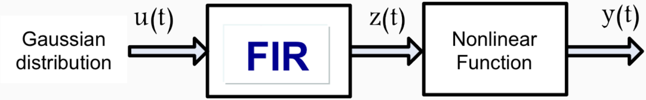
FIR Wiener model description

Why is the FIR approximation needed?



- $z(t)$ is not accessible for IIR estimation

FIR model (for now with Gaussian input assumption)



$$\varphi(t) \triangleq \begin{bmatrix} u(t) \\ u(t-1) \\ \vdots \\ u(t-n+1) \end{bmatrix}, \quad \theta \triangleq \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(n-1) \end{bmatrix}$$

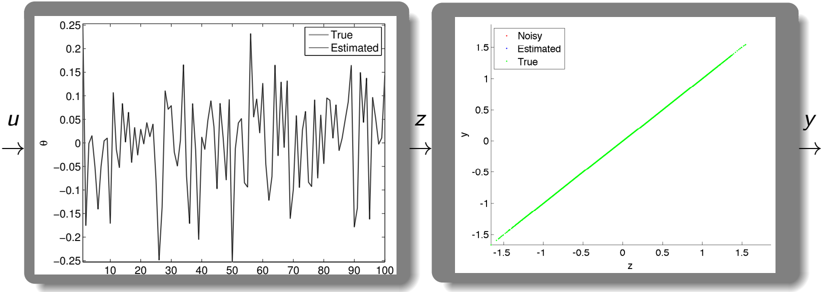
$$y(t) = f(\varphi^\top(t)\theta)) + e(t)$$

How to get the linear part only? Does the BLA work?

Outline

How does the BLA perform ?

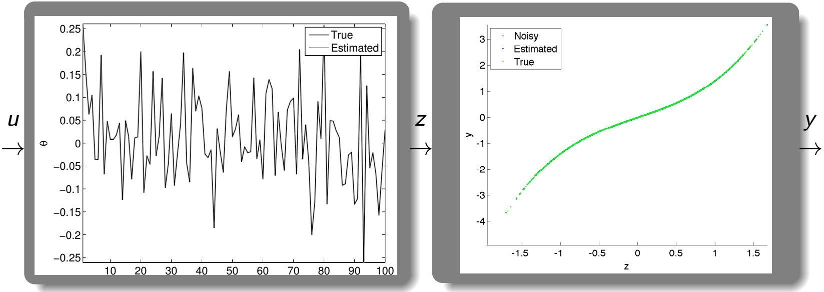
$N = 100000$, $n_\theta = 100, SNR = \infty$, nonlinearity : $y = z$;



- BLA=LS is at its best : optimal estimator
- Try some real Wiener system

How does the BLA perform ?

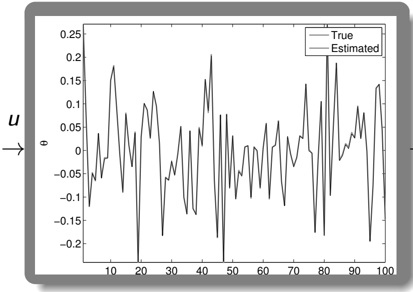
$N = 100000$, $n_\theta = 100, SNR = \infty$, nonlinearity : $y = z + 0.5z^3$;

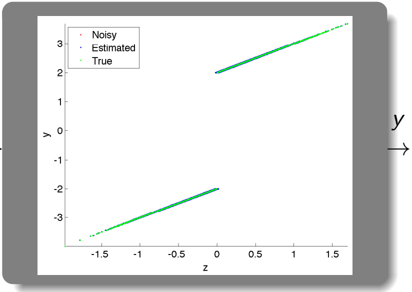


- BLA still works fine
- Maybe the nonlinearity is *"not nonlinear enough"*
- Try something more nonlinear

How does the BLA perform ?

$N = 100000$, $n_\theta = 100, SNR = \infty$, nonlinearity : $y = 2\text{sign}(z) + z$;



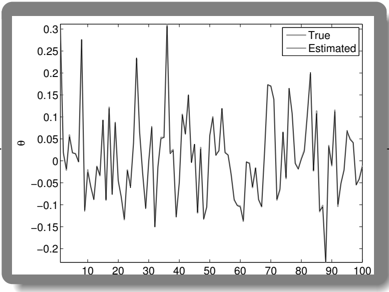


- BLA seems not to be disturbed by discontinuities

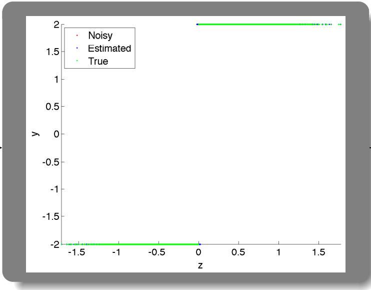
How does the BLA perform ?

$N = 100000$, $n_\theta = 100, SNR = \infty$, nonlinearity : $y = 2\text{sign}(z)$;

$\overset{u}{\rightarrow}$



$\overset{z}{\rightarrow}$

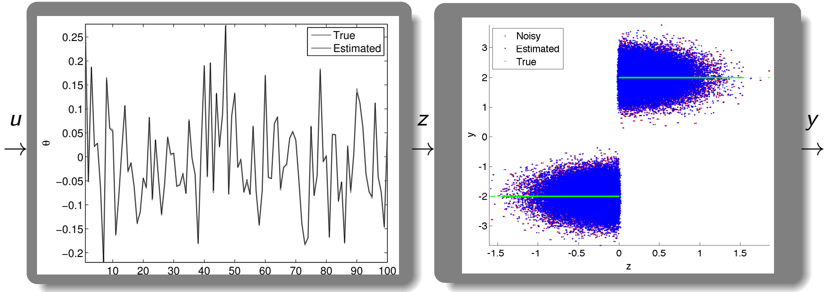


$\overset{y}{\rightarrow}$

- Not even for saturation case
- Try add some noise

How does the BLA perform ?

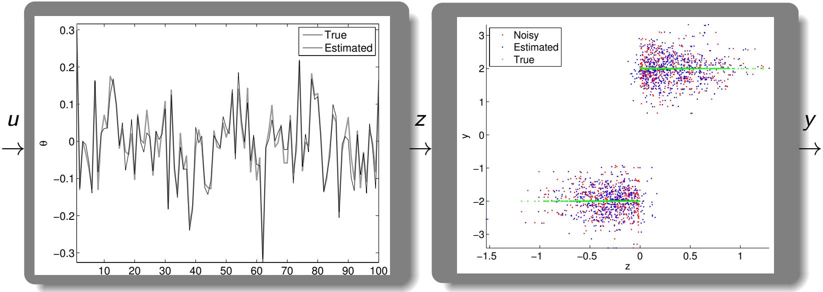
$N = 100000$, $n_\theta = 100$, $SNR = 10dB$, nonlinearity : $y = 2\mathrm{sign}(z)$;



- The linear part estimation is still really close to the truth
- Reduce the number of points

How does the BLA perform ?

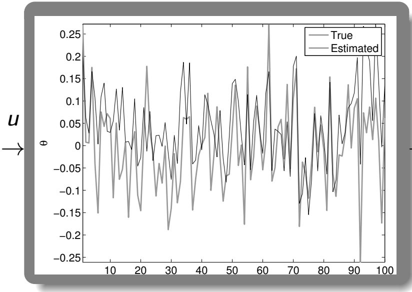
$N = 1000$, $n_\theta = 100, SNR = 10dB$, nonlinearity : $y = 2\mathrm{sign}(z)$;

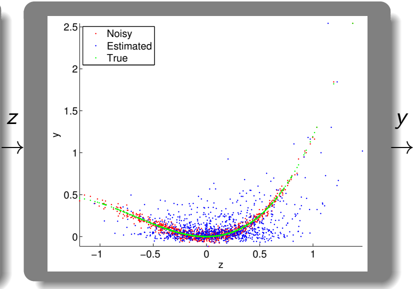


- The linear part estimation is still acceptable ($N = 10n$)
- What is the catch ?
- All functions were monotonous
- Try non-monotonous functions

How does the BLA perform ?

$N = 1000,\ n_\theta = 100, SNR = 10dB$, nonlinearity : $y = 0.8z^3 + 0.2z^2$;





- Now it is off !
- Though, asymptotically $\theta_{\text{BLA}} \to \theta_0$
- Is it possible to make the BLA fail ?
- Are Wiener identification efforts worth it ?

What does the BLA do

The LS from the PEM point of view

- $y(k) = \varphi(k)^\top \theta_o + e(k)$
- $\hat{\theta}_{LS} = \min_\theta ||e(k)||_2$
- $\hat{\theta}_{LS} = \mathbb{E}\left[\left(\varphi(k)^\top \varphi(k)\right)^{-1}\right] \mathbb{E}\left[\varphi(k)^\top y(k)\right]$

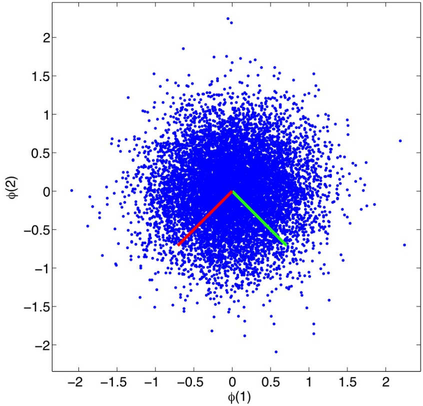
In the Wiener system case

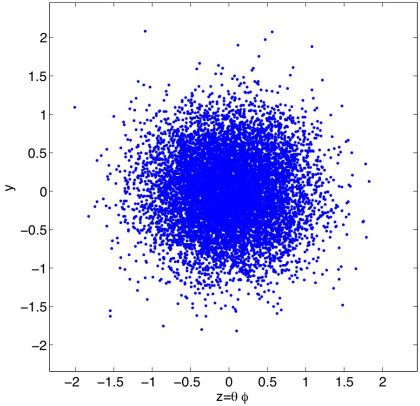
- $y(k) = f(\varphi(k)^\top \theta_o) + e(k)$
- In Wiener systems, we are **NOT** interested in θ_o
  - Wiener model is ill-posed
  - The output is the same for any pair $(\lambda\theta_o, f(z/\lambda)), \lambda \in \mathbb{R}$
  - Usual constraint : $||\theta||_2 = 1$
  - **Any $\hat{\theta} = \lambda\theta_o, \lambda \in \mathbb{R}$ is acceptable**$\rightarrow \hat{\theta} = \hat{\theta}/||\hat{\theta}||_2$

The Wiener system identification problem

The identification problem as a geometric interpretation

- Any $\lambda\theta_o$ is suitable $\rightarrow$ we are searching for a projection direction
- How to manipulate φ in order to find the correct projection ?

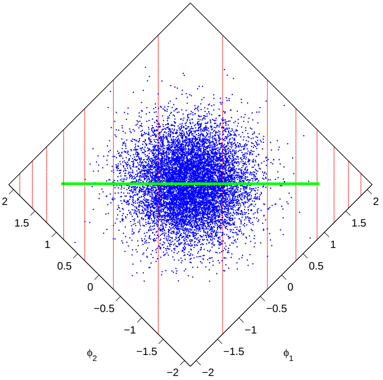




What does the BLA do ?

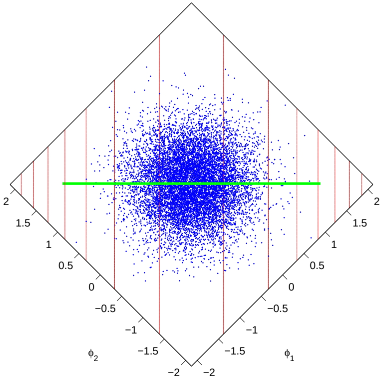
- $\hat{\theta}_{\text{BLA}} = \mathbb{E}\left[\left(\varphi(k)^\top \varphi(k)\right)^{-1}\right] \mathbb{E}\left[\varphi(k)^\top y(k)\right] = \lambda_1 \mathbb{E}\left[\varphi(k)^\top y(k)\right]$
- Where is y in the φ space ?

What does the BLA do ?



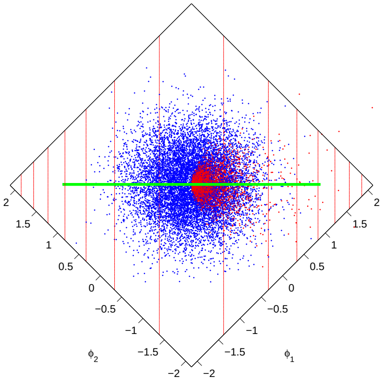
What does the BLA do ?

$\varphi(k)$



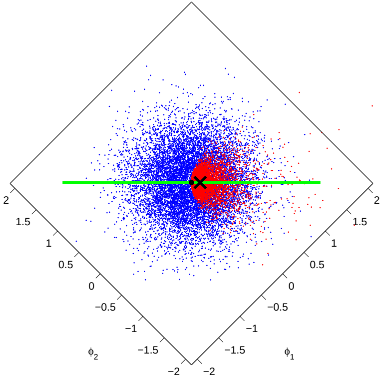
What does the BLA do ?

$\varphi(k)^\top y(k)$



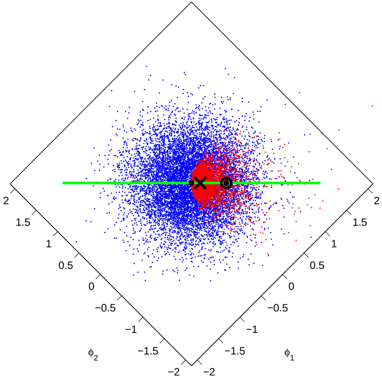
What does the BLA do ?

$$\mathbb{E}\left[\varphi(k)^\top y(k)\right]$$



What does the BLA do ?

$$\hat{\theta}_{\mathrm{BLA}} = \mathbb{E}\left[\left(\varphi(k)^\top \varphi(k)\right)^{-1}\right]\mathbb{E}\left[\varphi(k)^\top y(k)\right]$$



What does the BLA do ?

Conclusions

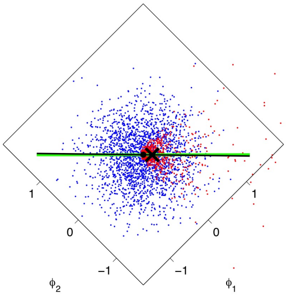
- BLA is surely not optimal
- It can be though proven : $\hat{\theta}_{\mathrm{BLA}} \to \lambda\theta_{\mathrm{o}}, \lambda \in \mathbb{R}$
- If enough data, the Wiener identification is then solved

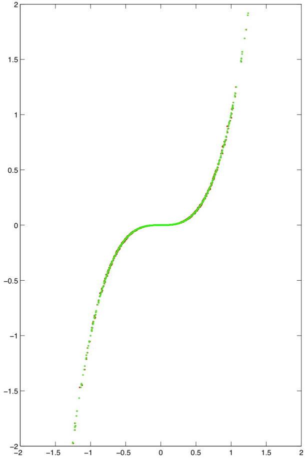
Not exactly

- $\hat{\theta}_{\mathrm{BLA}} = \underbrace{\mathbb{E}\left[\left(\varphi(k)^\top\varphi(k)\right)^{-1}\right]\mathbb{E}\left[z(k)f(z_k)\right]}_{\lambda}\theta_{\mathrm{o}}$
- If f is even, then $\mathbb{E}\left[z(k)f(z_k)\right] = 0$ and asymptotically $\hat{\theta}_{\mathrm{BLA}} \to 0$
- Starting statement for our work
- Should we really bother only for even functions ?

Yes, Life is not asymptotic

Example in 2D



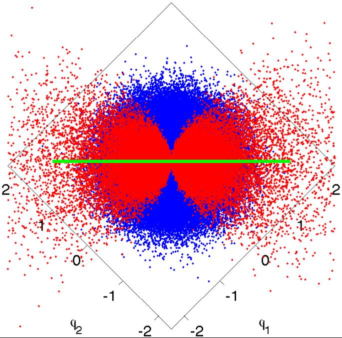


Outline

The BLA limitation

Why is the BLA limited ?

- For even nonlinear functions $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)\right] \rightarrow 0$
- Is there any other operator which would be robust to even nonlinearities ?
- *"Easy to see"* the *"main direction"* of $\varphi(k)^\top y(k)$ (red cloud)
- Make a estimator based on PCA$\rightarrow \mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right]$



The WPCA algorithm

The main proof

- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right] = \sigma_u^2 \mathbb{E}\left[y^2\right] I_n + \operatorname{cov}\left(z^2, y^2\right) \theta\theta^\top$

- $\operatorname{cov}\left(z^2, y^2\right)$ and $\sigma_u^2 \mathbb{E}\left[y^2\right]$ are real numbers
- $\sigma_u^2 \mathbb{E}\left[y^2\right] I_n$ has n eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$
- $\theta\theta^\top$ has $\lambda_1 = 1$ with associated eigenvector $\pm\theta$ and $n-1$ eigenvalues equal to 0
- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right]$ has :
  - 1 eigenvalue equal to $\sigma_u^2 \mathbb{E}\left[y^2\right] + \operatorname{cov}\left(z^2, y^2\right)$ with associated eigenvector $\pm\theta$
  - $n-1$ eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$

Proposed estimator

Choose $\hat{\theta}$ as the eigenvector associated to the "different" eigenvalue : smallest or largest depending on $\operatorname{sign}\left(\operatorname{cov}\left(z^2, y^2\right)\right)$

The WPCA algorithm

The main proof

- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right] = \sigma_u^2 \mathbb{E}\left[y^2\right] I_n + \operatorname{cov}\left(z^2, y^2\right)\theta\theta^\top$

- $\operatorname{cov}\left(z^2, y^2\right)$ and $\sigma_u^2 \mathbb{E}\left[y^2\right]$ are real numbers
- $\sigma_u^2 \mathbb{E}\left[y^2\right] I_n$ has n eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$
- $\theta\theta^\top$ has $\lambda_1 = 1$ with associated eigenvector $\pm\theta$ and $n-1$ eigenvalues equal to 0
- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right]$ has :
  - 1 eigenvalue equal to $\sigma_u^2 \mathbb{E}\left[y^2\right] + \operatorname{cov}\left(z^2, y^2\right)$ with associated eigenvector $\pm\theta$
  - $n-1$ eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$

Proposed estimator

Choose $\hat{\theta}$ as the eigenvector associated to the "different" eigenvalue : smallest or largest depending on $\operatorname{sign}\left(\operatorname{cov}\left(z^2, y^2\right)\right)$

The WPCA algorithm

The main proof

- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right] = \sigma_u^2 \mathbb{E}\left[y^2\right] I_n + \mathrm{cov}\left(z^2, y^2\right) \theta\theta^\top$

- $\mathrm{cov}\left(z^2, y^2\right)$ and $\sigma_u^2 \mathbb{E}\left[y^2\right]$ are real numbers
- $\sigma_u^2 \mathbb{E}\left[y^2\right] I_n$ has n eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$
- $\theta\theta^\top$ has $\lambda_1 = 1$ with associated eigenvector $\pm\theta$ and $n-1$ eigenvalues equal to 0
- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right]$ has :
  - 1 eigenvalue equal to $\sigma_u^2 \mathbb{E}\left[y^2\right] + \mathrm{cov}\left(z^2, y^2\right)$ with associated eigenvector $\pm\theta$
  - $n-1$ eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$

Proposed estimator

Choose $\hat{\theta}$ as the eigenvector associated to the "different" eigenvalue : smallest or largest depending on $\mathrm{sign}\left(\mathrm{cov}\left(z^2, y^2\right)\right)$

The WPCA algorithm

The main proof

- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right] = \sigma_u^2 \mathbb{E}\left[y^2\right] I_n + \operatorname{cov}\left(z^2, y^2\right) \theta\theta^\top$

- $\operatorname{cov}\left(z^2, y^2\right)$ and $\sigma_u^2 \mathbb{E}\left[y^2\right]$ are real numbers
- $\sigma_u^2 \mathbb{E}\left[y^2\right] I_n$ has n eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$
- $\theta\theta^\top$ has $\lambda_1 = 1$ with associated eigenvector $\pm\theta$ and $n-1$ eigenvalues equal to 0
- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right]$ has :
  - 1 eigenvalue equal to $\sigma_u^2 \mathbb{E}\left[y^2\right] + \operatorname{cov}\left(z^2, y^2\right)$ with associated eigenvector $\pm\theta$
  - $n-1$ eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$

Proposed estimator

Choose $\hat{\theta}$ as the eigenvector associated to the "different" eigenvalue : smallest or largest depending on $\operatorname{sign}\left(\operatorname{cov}\left(z^2, y^2\right)\right)$

The WPCA algorithm

The main proof

- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right] = \sigma_u^2 \mathbb{E}\left[y^2\right] I_n + \mathrm{cov}\left(z^2, y^2\right) \theta\theta^\top$

- $\mathrm{cov}\left(z^2, y^2\right)$ and $\sigma_u^2 \mathbb{E}\left[y^2\right]$ are real numbers
- $\sigma_u^2 \mathbb{E}\left[y^2\right] I_n$ has n eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$
- $\theta\theta^\top$ has $\lambda_1 = 1$ with associated eigenvector $\pm\theta$ and $n-1$ eigenvalues equal to 0
- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top \left(\varphi(k)^\top y(k)\right)\right]$ has :
  - 1 eigenvalue equal to $\sigma_u^2 \mathbb{E}\left[y^2\right] + \mathrm{cov}\left(z^2, y^2\right)$ with associated eigenvector $\pm\theta$
  - $n-1$ eigenvalues equal to $\sigma_u^2 \mathbb{E}\left[y^2\right]$

Proposed estimator

Choose $\hat{\theta}$ as the eigenvector associated to the "different" eigenvalue : smallest or largest depending on $\mathrm{sign}\left(\mathrm{cov}\left(z^2, y^2\right)\right)$

The WPCA algorithm

The main proof

- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top\left(\varphi(k)^\top y(k)\right)\right] = \sigma_u^2\mathbb{E}\left[y^2\right] I_n + \operatorname{cov}\left(z^2, y^2\right)\theta\theta^\top$

- $\operatorname{cov}\left(z^2, y^2\right)$ and $\sigma_u^2\mathbb{E}\left[y^2\right]$ are real numbers
- $\sigma_u^2\mathbb{E}\left[y^2\right] I_n$ has n eigenvalues equal to $\sigma_u^2\mathbb{E}\left[y^2\right]$
- $\theta\theta^\top$ has $\lambda_1 = 1$ with associated eigenvector $\pm\theta$ and $n-1$ eigenvalues equal to 0
- $\mathbb{E}\left[\left(\varphi(k)^\top y(k)\right)^\top\left(\varphi(k)^\top y(k)\right)\right]$ has :
  - 1 eigenvalue equal to $\sigma_u^2\mathbb{E}\left[y^2\right] + \operatorname{cov}\left(z^2, y^2\right)$ with associated eigenvector $\pm\theta$
  - $n-1$ eigenvalues equal to $\sigma_u^2\mathbb{E}\left[y^2\right]$

Proposed estimator

Choose $\hat{\theta}$ as the eigenvector associated to the "different" eigenvalue : smallest or largest depending on $\operatorname{sign}\left(\operatorname{cov}\left(z^2, y^2\right)\right)$

Slight details

The proof holds for any function of y : $w(y(k)) = w(k)$

- $\mathbb{E}\left[\left(\varphi(k)^\top w(k)\right)^\top \left(\varphi(k)^\top w(k)\right)\right] = \sigma_u^2 \mathbb{E}\left[w^2\right] I_n + \operatorname{cov}\left(z^2, w^2\right) \theta\theta^\top$
- Variance can be reduced *e.g.* : $w(y) = \sqrt{|y|}, 1 - e^{\alpha y}$. Empirically $\ln(1 + |y|)$.

$\varphi(k)$ normalization

- In the BLA, there is a normalization by $\mathbb{E}(\varphi(k)^\top \varphi(k))$
- Here, the proof holds asymptotically for gaussian assumption
- In practice $\varphi(t)$ needs to be normalized such there isn't any *a priori* predominating eigenvalue in φ :

$$\Phi = \begin{bmatrix} \varphi^\top(1) \\ \varphi^\top(2) \\ \vdots \\ \varphi^\top(N) \end{bmatrix}, \quad Y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \quad \rightarrow \Phi = USV^\top \quad \rightarrow \tilde{\Phi} \triangleq \begin{bmatrix} \tilde{\varphi}(1) \\ \tilde{\varphi}(2) \\ \vdots \\ \tilde{\varphi}(N) \end{bmatrix} := UV^\top$$

The algorithm

Step 1 : Normalization

- From $\Phi = USV^\top$, compute $\tilde{\Phi} = UV^\top$

Step 2 : Weighting

$$\tilde{\Phi}_w := \begin{bmatrix} w(y(1))\tilde{\varphi}(1) \\ w(y(2))\tilde{\varphi}(2) \\ \vdots \\ w(y(N))\tilde{\varphi}(N) \end{bmatrix}$$

Step 3 : SVD

- $\tilde{\Phi}_w = U_w S_w V_w^\top$ and set $\tilde{\theta} = V_w^{(1)}$ or $\tilde{\theta} = V_w^{(n)}$ (the one farthest from median)

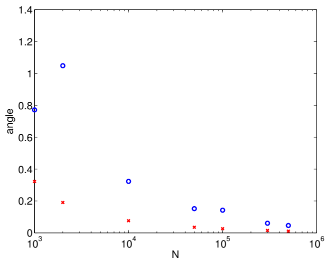
Step 4 : "Denormalization"

- $\theta = VS^{-1}V^\top\tilde{\theta} \rightarrow \theta = \dfrac{\theta}{||\theta||}$

Some programming aspects

Step 1 : SVD Nxn

- n= 100 , N=3000 $\rightarrow t = 0.08s$
- n= 100 , N=30000 $\rightarrow t = 0.4s$
- n= 100 , N=300000 $\rightarrow t = 5s$
- n= 100 , N=500000 $\rightarrow t = 10s$
- n= 100 , N=600000 $\rightarrow t = X$



Step 1 reduces the variance : for large datasets, Step 1 can be skipped

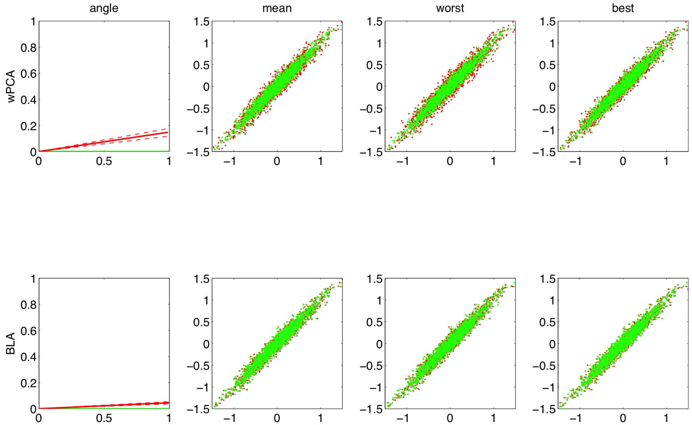
Step 3 : SVD nxn

- n= 100 $\rightarrow t \approx t_{BLA} < 0.1s$

- Very simple algorithm
- Similar computation time as the BLA

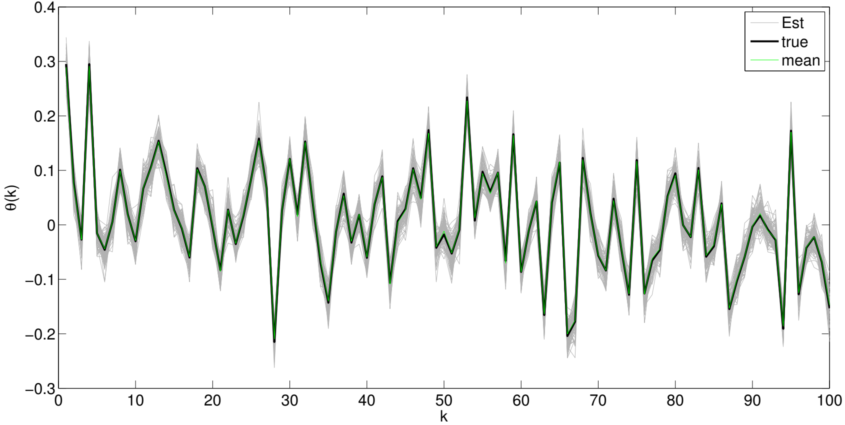
Outline

Example : n=100, N=2000, SNR=12dB, $w = \ln(1 + |y|)$



Example : n=100, N=2000, SNR=12dB, $w = \ln(1 + |y|)$

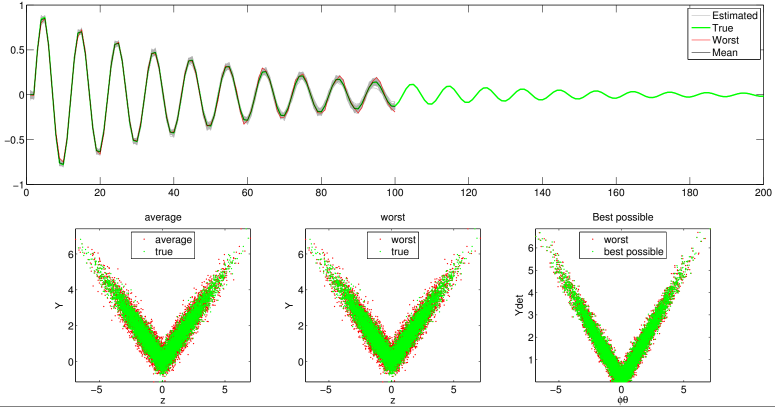
In the θ space for a median angle :



Example : n=100, N=10000, SNR=10dB

With a transfer function

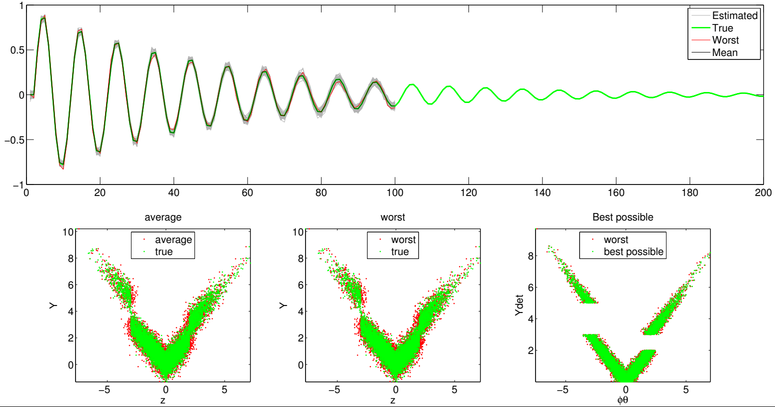
- $(1 - 1.69q^{-1} + 1.119q^{-2} - 0.096q^{-3})z(k) = 0.5q^{-3}u(k)$
- $f(z) = |z|,\ w(y) = ln(1 + |y|)$



Example : n=100, N=10000, SNR=10dB

With a transfer function

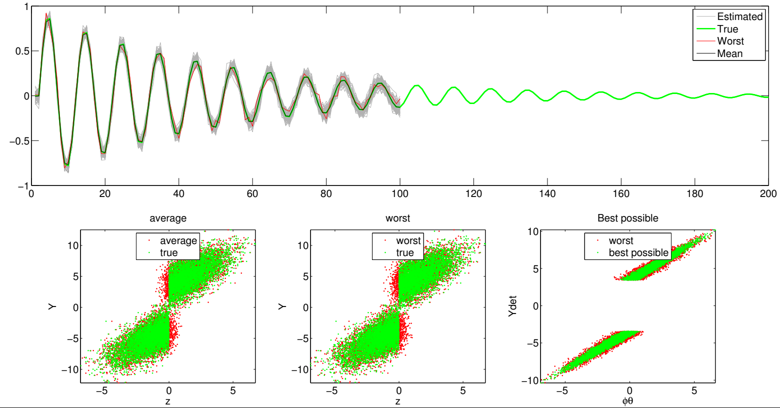
- $(1 - 1.69q^{-1} + 1.119q^{-2} - 0.096q^{-3})z(k) = 0.5q^{-3}u(k)$
- $f(z) = f(|z|)$, $w(y) = ln(1 + |y|)$



Example : n=100, N=10000, SNR=10dB

With a transfer function

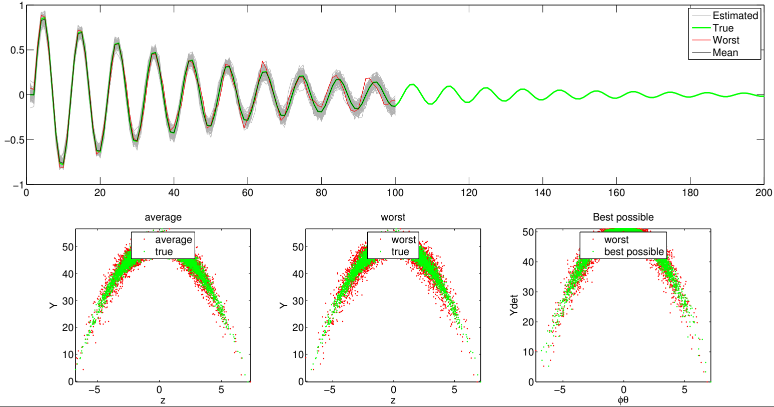
- $(1 - 1.69q^{-1} + 1.119q^{-2} - 0.096q^{-3})z(k) = 0.5q^{-3}u(k)$
- $f(z) = 0.5sign(z) + z,\ w(y) = ln(1 + |y|)$



Example : n=100, N=10000, SNR=10dB

With a transfer function

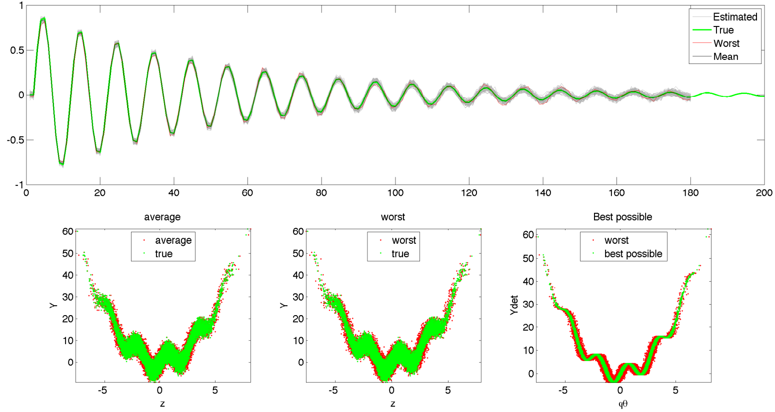
- $(1 - 1.69q^{-1} + 1.119q^{-2} - 0.096q^{-3})z(k) = 0.5q^{-3}u(k)$
- $f(z) = max(z) - z.^2$, $w(y) = ln(1 + |y|)$



Example : n=180, N=40000, SNR=10dB

With a transfer function

- $(1 - 1.69q^{-1} + 1.119q^{-2} - 0.096q^{-3})z(k) = 0.5q^{-3}u(k)$
- $f(z) = z^2 + 4\sin(6z\pi/max(|z|)),\ w(y) = ln(1 + |y|)$



Outline

Gaussian Hypothesis relaxation

What we are interested in

- $\mathbb{E}\left[\left(\varphi(k)^\top w(k)\right)^\top \left(\varphi(k)^\top w(k)\right)\right]$
- $\mathbb{E}\left[\phi(u)\right]$ with $u \sim g$, g is Gaussian PDF : $\mathbb{E}_g\left[\phi(u)\right]$

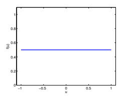
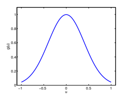
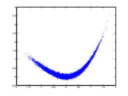
What we have

- $u \sim f$, (f not gaussian PDF), $\rightarrow \mathbb{E}_f\left[\phi(u)\right] \approx \frac{1}{N}\sum_{i=1}^{N} \phi_i(u)$
- Is it possible to get $\mathbb{E}_g\left[\phi(u)\right]$?

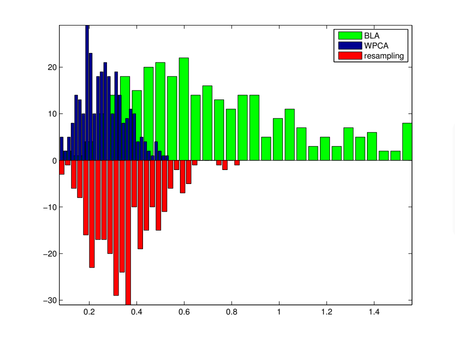
Importance sampling

- $\mathbb{E}_g\left[\phi(u)\right] = \int \phi(u)g(u)du = \int \phi(u)\frac{g(u)}{f(u)}f(u)du = \mathbb{E}_f\left[\phi(u)\frac{g(u)}{f(u)}\right]$
- $\mathbb{E}_g\left[\phi(u)\right] \approx \frac{1}{N}\sum_{i=1}^{N} \phi_i(u)\frac{g(u)}{f(u)}$, with $u \sim f$
- It is possible to estimate $\mathbb{E}_g\left[\phi(u)\right]$ from "any" f PDF distribution ! (asymptotically)

Example

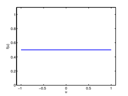
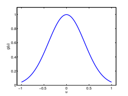
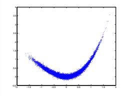
From  To  with NL= 

n_θ=6, N= 100, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

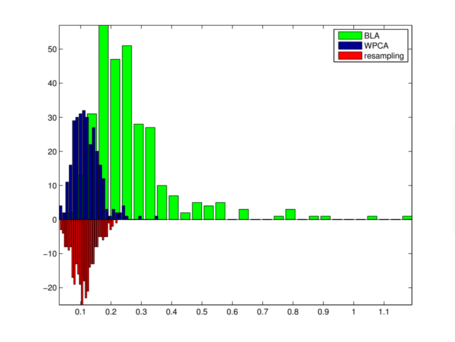


- BLA is seriously handicaped
- original and resampled distribution perform similarly

Example

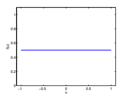
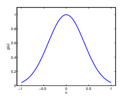
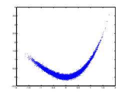
From  To  with NL= 

n_θ=6, N= 1000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

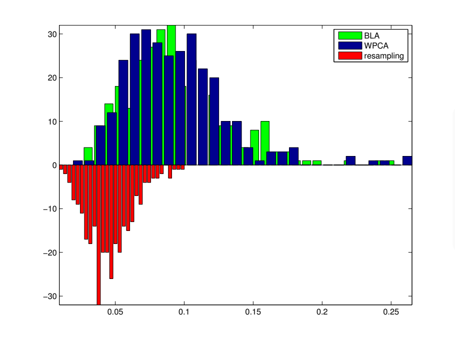


- BLA is seriously handicaped
- original and resampled distribution perform similarly

Example

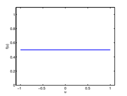
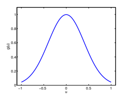
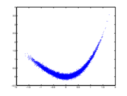
From  To  with NL= 

n_θ=6, N= 10000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

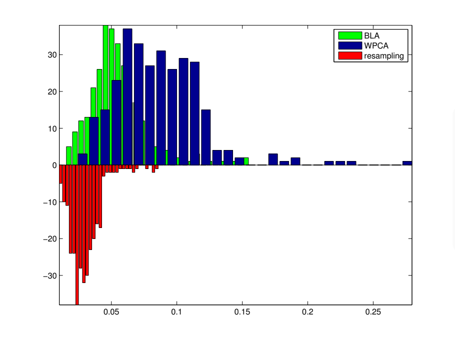


- BLA is advantaged by the uniform distribution
- Resampled distribution performs better than original

Example

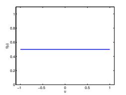
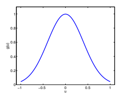
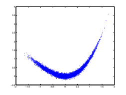
From  To  with NL= 

n_θ=6, N= 100000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$



- BLA is advantaged by the uniform distribution
- Resampled distribution performs better than original

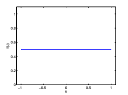
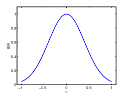
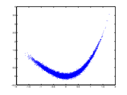
Example

From  To  with NL= 

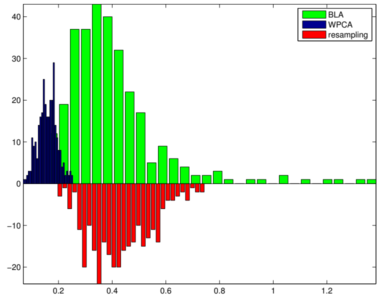
n_θ=6, N= 100000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

- Is the problem solved ?
- No free lunch : resampling has a cost
- Let us raise the number of dimension 6 $\rightarrow$ 12

Example

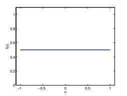
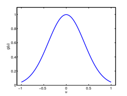
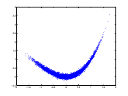
From  To  with NL= 

n_θ=12, N= 1000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

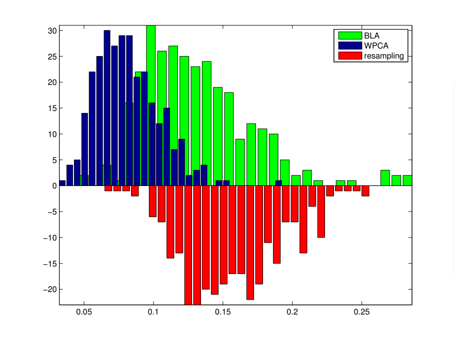


- BLA is seriously handicaped
- Resampled distribution performs much worse than original

Example

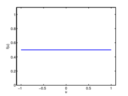
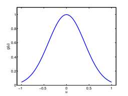
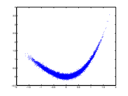
From  To  with NL= 

n_θ=12, N= 10000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

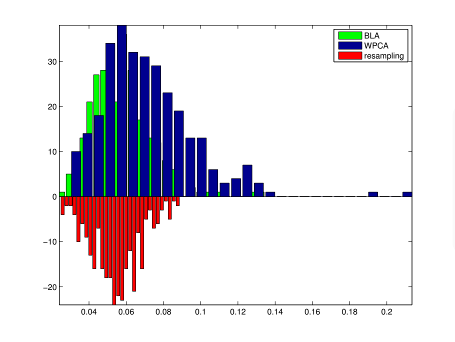


- BLA starts reducing the variance thanks to Uniform distribution
- Resampled distribution performs much worse than BLA

Example

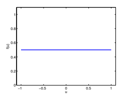
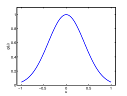
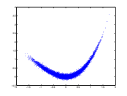
From  To  with NL= 

n_θ=12, N= 100000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$



- BLA tends to asymptotic results
- Resampled distribution does not perform better than BLA

Example

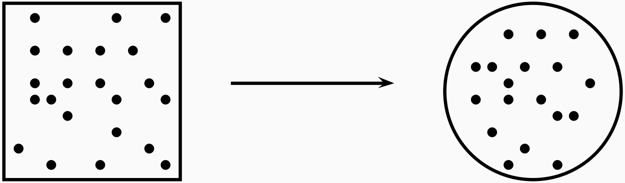
From  To  with NL= 

n_θ=12, N= 100000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$

- What is the cost of resampling ?

Extreme Example

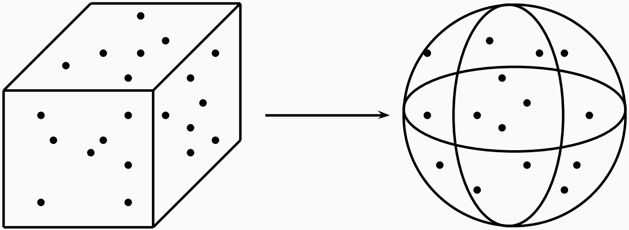
Consider an extreme case : Uniform to Circle PDF 2D



- Originally N points
- Finally $\pi 0.5^2 N = 0.79N$

Extreme Example

Consider an extreme case : Uniform to Circle PDF 3D



- Originally N points
- Finally $4/3\pi 0.5^3 N \approx 0.52N$

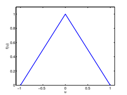
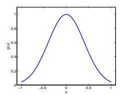
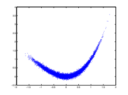
Extreme Example

Consider an extreme case : Uniform to Circle PDF more D

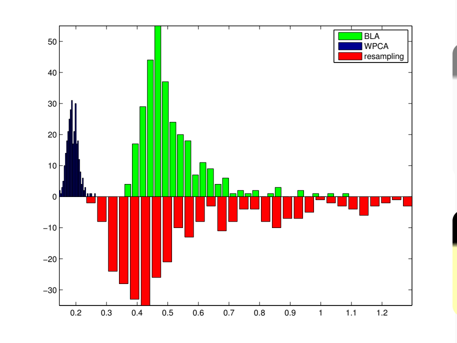
- $n_\theta = 5$
  - Originally N points
  - Finally $N \approx 0.16N$
- $n_\theta = 10$
  - Originally N points
  - Finally $N \approx 0.003N$
- $n_\theta = 20$
  - Originally N points
  - Finally $N \approx 0N$

- There is some curse of dimensionality
- **Resampling from one distribution to another = loss of information**

The good news and the bad news

From  To  with NL= 

n_θ=100, N= 10000, SNR=20dB, $f(z) = 0.2z^3 + 0.8z^2$



The bad news

Importance sampling will not perform well for high dimensional problems

The good news

WPCA approach works fine without resampling

Outline

Conclusions

The correlation-based methods

- Most identification methods
- Will always have trouble with parity components

Another view on identification

- Geometric interpretation
- WPCA based approaches

Facts

- The proposed approach is not optimal
  - what is optimality in approximation context ?
  - how to initialise optimal approaches ?

Conclusions

Advantages of the method

- It does not require usual *a priori* NL knowledge :
  - No need for invertibility, parity,continuity, monotony or fix point information
- It has a low computational effort
- It is robust to noise
- The gaussian assumption is not required in practice

Further work

- Better formulation for IIR linear blocks
- How to best choose the weighting function

Wiener System Identification by Weighted Principal Component Analysis

Q. Zhang **V. Laurain**

14 juin 2013
vincent.laurain@univ-lorraine.fr