TECHNISCHE UNIVERSITÄT DRESDEN

ComNets
Deutsche Telekom Chair of Communication Networks

Deutsche Telekom Chair for Communication Networks

Fast Source Separation

Study Thesis

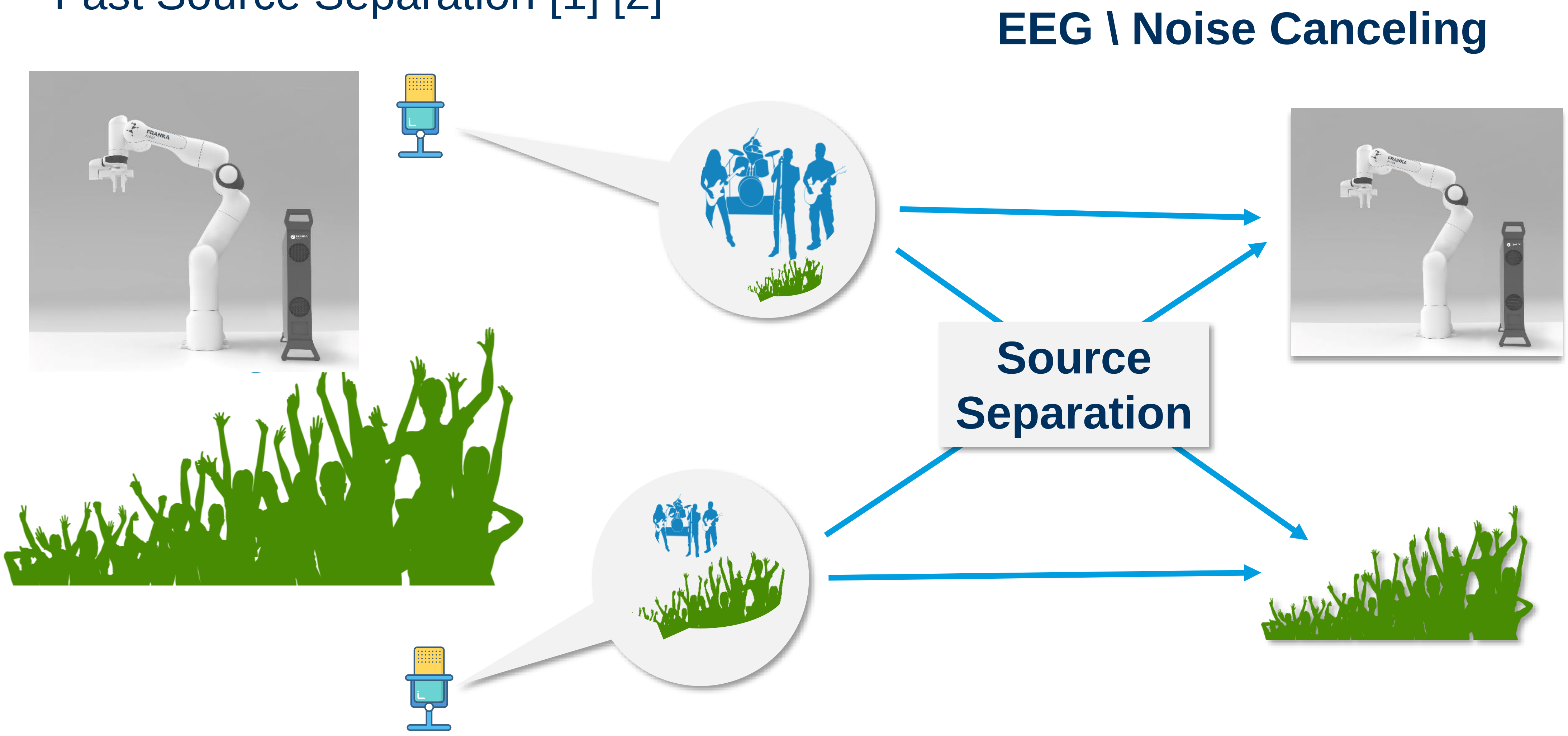
Yunbin Shen

Supervisors: M.Sc. Huanzhuo Wu

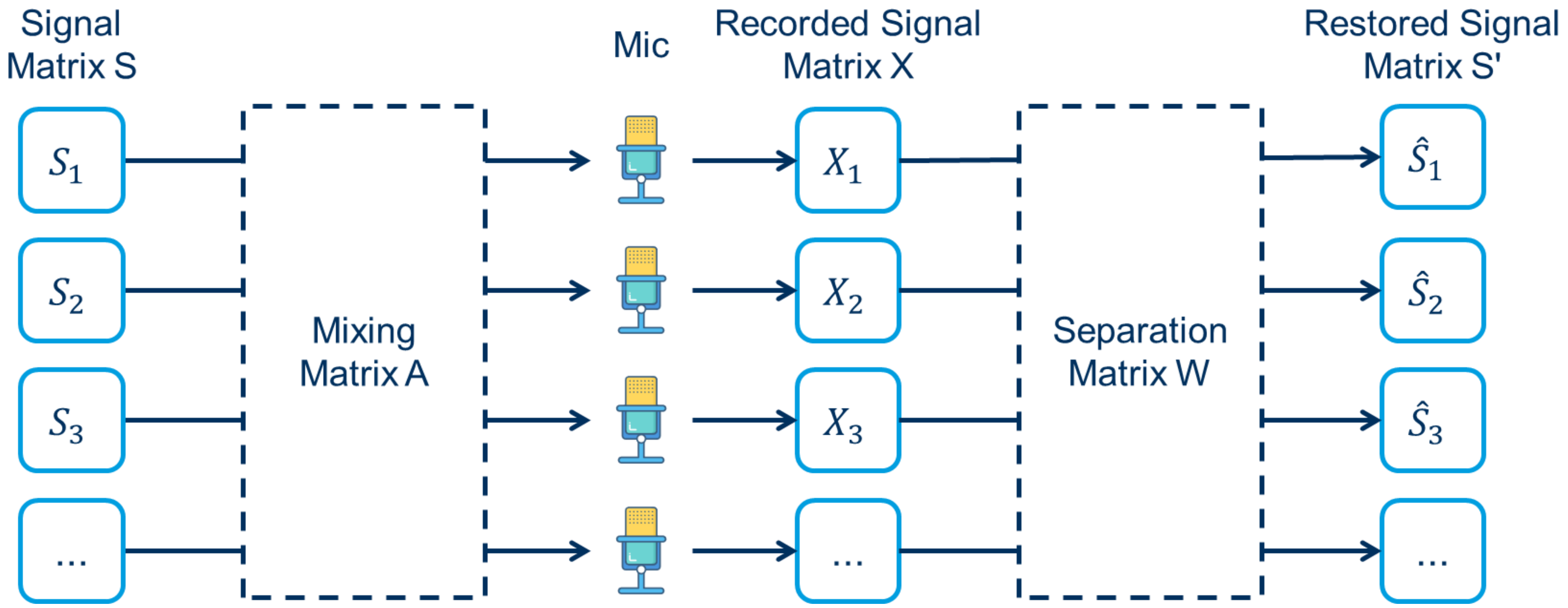
Dresden, 29.11.2019

Contents
Fast Source Separation

1 Introduction
Fast Source Separation [1] [2]



2 Problem Statement
Blind Source Separation



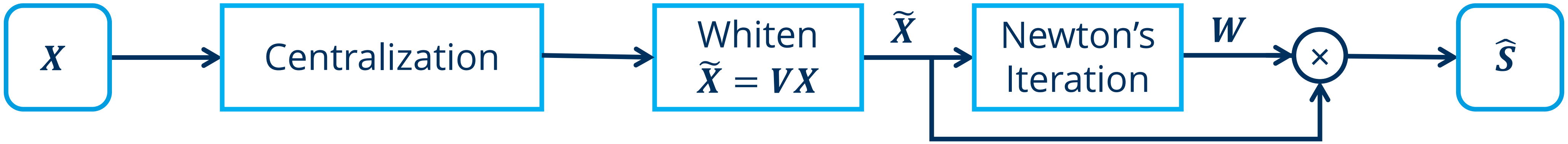
$$\hat{\boldsymbol{S}} = \boldsymbol{W}\boldsymbol{X} \approx \boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{A}^{-1}\boldsymbol{X} \quad (1)$$

2 Problem Statement
FastICA Algorithm [3] [8]

$$\boldsymbol{W}^{+} = \boldsymbol{E}\{\widetilde{\boldsymbol{X}}\boldsymbol{g}(\boldsymbol{W}^{T}\widetilde{\boldsymbol{X}})\} - \boldsymbol{E}\{\boldsymbol{g}'(\boldsymbol{W}^{T}\widetilde{\boldsymbol{X}})\}\boldsymbol{W}$$

$$\boldsymbol{W}^{+} = \boldsymbol{W}^{+}/\|\boldsymbol{W}^{+}\|$$

$$repeat\ until: Diff = Max\{\boldsymbol{W}^{+}\boldsymbol{W}^{T} - \boldsymbol{I}\} < Tol$$



8 * 4000ms recorded audio, f_s=16000Hz, i5 2.6Ghz, Win 10, Python3.7

Slow! >232ms

FastICA Time Line

Whiten 12ms	Newton's Iteration 220-470ms

2 Problem Statement
FastICA Algorithm [3] [8]

$$W^{+} = E\{\tilde{X}g(W^{T}\tilde{X})\} - E\{g'(W^{T}\tilde{X})\}W$$
$$W^{+} = W^{+}/||W^{+}||$$

$$repeat\ until: Diff = Max\{W^{+}W^{T} - I\} < Tol$$

$$1, Random\ W_{0}^{n\times n}$$

$$When\ W_{0_{n\times n}} \approx W_{n\times n} \implies Iteration\ Number \downarrow$$

$$2, Mixed\ signals\ \tilde{X}^{n\times m}$$

$$When\ \tilde{X}_{n\times m} \rightarrow \tilde{X}_{n\times k}, k < m \implies Computation \downarrow$$

Newton's Iteration ↑

Usually: change the convergence speed (**2X Speed**)

eighth-order Newton's Iteration [4]

finding the optimal iterative step length [5]

using a special overrelaxation factor in g'(X) [6]

5-8X Speed !

1. Initial Separation Matrix Estimation

2. Multi-Level Newton's Iteration

3 Methodology
Initial Separation Matrix Estimation

From Eq.(1), we have

$$E(\boldsymbol{x}_i) = \sum_{l=1}^{n} a_{i,l} E(\boldsymbol{s}_l) \qquad (2)$$

$$\frac{E(\boldsymbol{x}_i)}{E(\boldsymbol{x}_j)} = \frac{\sum_{l=1}^{n} a_{i,l} E(\boldsymbol{s}_l)}{\sum_{l=1}^{n} a_{j,l} E(\boldsymbol{s}_l)} = \frac{a_{i,1}E(\boldsymbol{s}_1) + a_{i,2}E(\boldsymbol{s}_2) + \cdots + a_{i,n}E(\boldsymbol{s}_n)}{a_{j,1}E(\boldsymbol{s}_1) + a_{j,2}E(\boldsymbol{s}_2) + \cdots + a_{j,n}E(\boldsymbol{s}_n)} \qquad (3)$$

When

$$\boxed{\begin{aligned} |E(\boldsymbol{s}_k)| &\gg \left| \sum_{l=1, l\neq k}^{n} a_{i,l} E(\boldsymbol{s}_l) \right| \\ |E(\boldsymbol{s}_k)| &\gg \left| \sum_{l=1, l\neq k}^{n} a_{j,l} E(\boldsymbol{s}_l) \right| \end{aligned} \qquad (4)}$$

Then we have,

$$\frac{a_{i,k}}{a_{j,k}} \approx \frac{E(\boldsymbol{x}_i)}{E(\boldsymbol{x}_j)} \qquad (5)$$

Proof in Appendix 1

3 Methodology
Initial Separation Matrix Estimation

From Eq.(12), we have

$$\widehat{\boldsymbol{A}} = \begin{bmatrix} 1 & \cdots & \hat{a}_{1,n} \\ \hat{a}_{2,1} & \cdots & \hat{a}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{a}_{n,1} & \cdots & 1 \end{bmatrix} \approx \begin{bmatrix} \frac{a_{1,1}}{a_{1,1}} & \cdots & \frac{a_{1,n}}{a_{n,n}} \\ \frac{a_{2,1}}{a_{1,1}} & \cdots & \frac{a_{1,n}}{a_{n,n}} \\ \vdots & \ddots & \vdots \\ \frac{a_{n,1}}{a_{1,1}} & \cdots & \frac{a_{n,n}}{a_{n,n}} \end{bmatrix} = \boldsymbol{A} \times \begin{bmatrix} \frac{1}{a_{1,1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{a_{n,n}} \end{bmatrix} = \boldsymbol{AK} \quad (6)$$

So that, we can get a separation matrix from $\widehat{\boldsymbol{A}}$

$$\boxed{\widehat{\boldsymbol{W}}} = \widehat{\boldsymbol{A}}^{-1} \approx (\boldsymbol{AK})^{-1} = \boldsymbol{K}^{-1}\boldsymbol{A}^{-1} = \boldsymbol{K}^{-1}\boldsymbol{W} \quad (7)$$

$$\boxed{Estimated\ intial\ \widehat{\boldsymbol{W}}_0{}^{n\times n} = \frac{\widehat{\boldsymbol{W}}}{\|\widehat{\boldsymbol{W}}\|}}$$

Proof in Appendix 2

3 Methodology
Initial Sealing Matrix Estimation

$\widehat{\boldsymbol{A}}$ can be computed as follows, where b_{l_i} is the positions of extracted time slots

$$\frac{E\left(x_{k,b_{l_i}}\right)}{E\left(x_{l,b_{l_i}}\right)} = \hat{a}_{i,l} \approx \frac{a_{i,l}}{a_{l,l}}$$
$$s.t.\ \boldsymbol{B}_l = \left[b_{l_1}, b_{l_2}, b_{l_3}, \ldots, b_{l_{t_l}}\right], \forall x_{l,b_{l_i}} > \left|x_{k,b_{l_i}}\right| \quad (8)$$
$$l = 1, \cdots, n \quad k = 1, \cdots, n \quad k \neq l$$

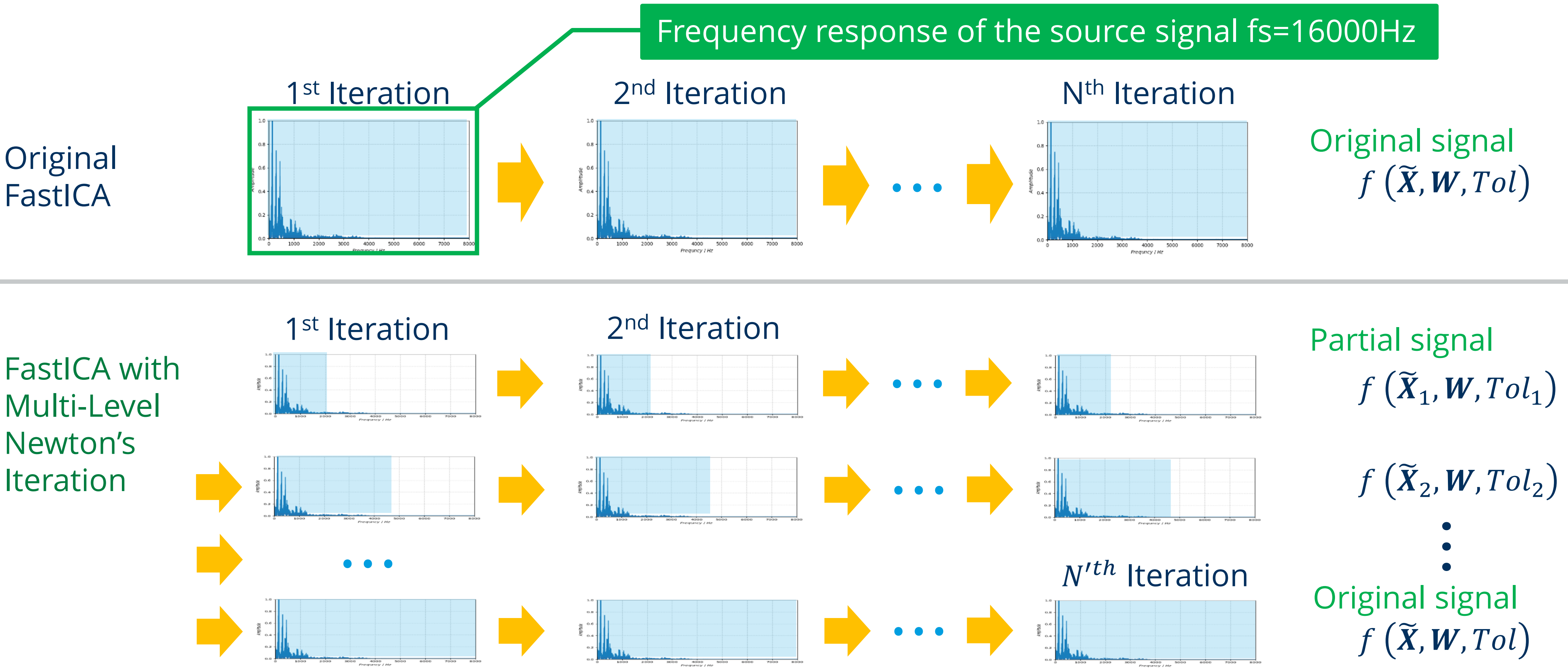
Proof in Appendix 1

The whole Algorithm is,

$$\widehat{\boldsymbol{A}} = \begin{bmatrix} 1 & \cdots & \hat{a}_{1,n} \\ \hat{a}_{2,1} & \cdots & \hat{a}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{a}_{n,1} & \cdots & 1 \end{bmatrix} \Rightarrow \widehat{\boldsymbol{W}} = \widehat{\boldsymbol{A}}^{-1} \Rightarrow Estimated\ \widehat{\boldsymbol{W}}_0{}^{n\times n} = \frac{\widehat{\boldsymbol{W}}}{\|\widehat{\boldsymbol{W}}\|}$$

Proof in Appendix 2

3 Methodology
Multi-Level Newton's Iteration



How to extract? Tolerance?

detail in Appendix 3

3 Methodology
Multi-Level Newton's Iteration

- Data extraction, with fixed interval e

$$ {}_e^i\boldsymbol{X} = [x_i \quad x_{i+e} \quad x_{i+2e} \quad \cdots \quad x_{i+e\lfloor n/e-1 \rfloor}] \quad (9) $$

Equivalent to low frequency sampling

The difference of $\boldsymbol{W}^+\boldsymbol{W}^T$ is defined as

$$ Diff = Max\{\boldsymbol{W}^+\boldsymbol{W}^T - \boldsymbol{I}\} < Tol \quad (10) $$

Because a bigger extraction interval e will leads to worse accuracy $\boldsymbol{W}$. Therefore, the tolerance should increase as e grows. Otherwise, a small tolerance will spend a lot of time.

- The tolerance of convergence for extracted signals can be determined as,

$$ Tol_{{}_e^i\tilde{\mathbf{X}}} \approx \sqrt{e} \cdot Tol_{\tilde{\mathbf{X}}} \quad (11) $$

Proof in Appendix 4

Tolerance Changed!

- The next extraction interval e is the maximum value of e under the condition $Diff > Tol_{{}_e^i\tilde{\mathbf{X}}}$.

3 Methodology
Ultra-FastICA

FastICA

X → Centraliz-ation → Whiten → Random W → Newton's Iteration → $\hat{S}$ old

Improved FastICA with Initial Separation Matrix Estimation

X → Centraliz-ation → Whiten → Estimated W → Newton's Iteration → $\hat{S}$ new

Improved FastICA with Multi-Level Newton's Iteration

X → Centraliz-ation → Whiten → Random W → Multi-Level Newton's Iteration → $\hat{S}$ new

Ultra-FastICA*

X → Centraliz-ation → Whiten → Estimated W → Multi-Level Newton's Iteration → $\hat{S}$ new

4 Evaluation
Setup & Testing Data

Sources:

— 63 human voices

Mixing matrix:

— $$\hat{A} = \begin{bmatrix} 1 & \cdots & \hat{a}_{1,n} \\ \hat{a}_{2,1} & \cdots & \hat{a}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{a}_{n,1} & \cdots & 1 \end{bmatrix} \quad \hat{a}_{\mathrm{i},j} \in \mathcal{N}\left(\mu = 0, 0 < \sigma^2 < 0.11\right) \ with \ \Lambda = \frac{\sigma}{0.33} \times 100\%$$

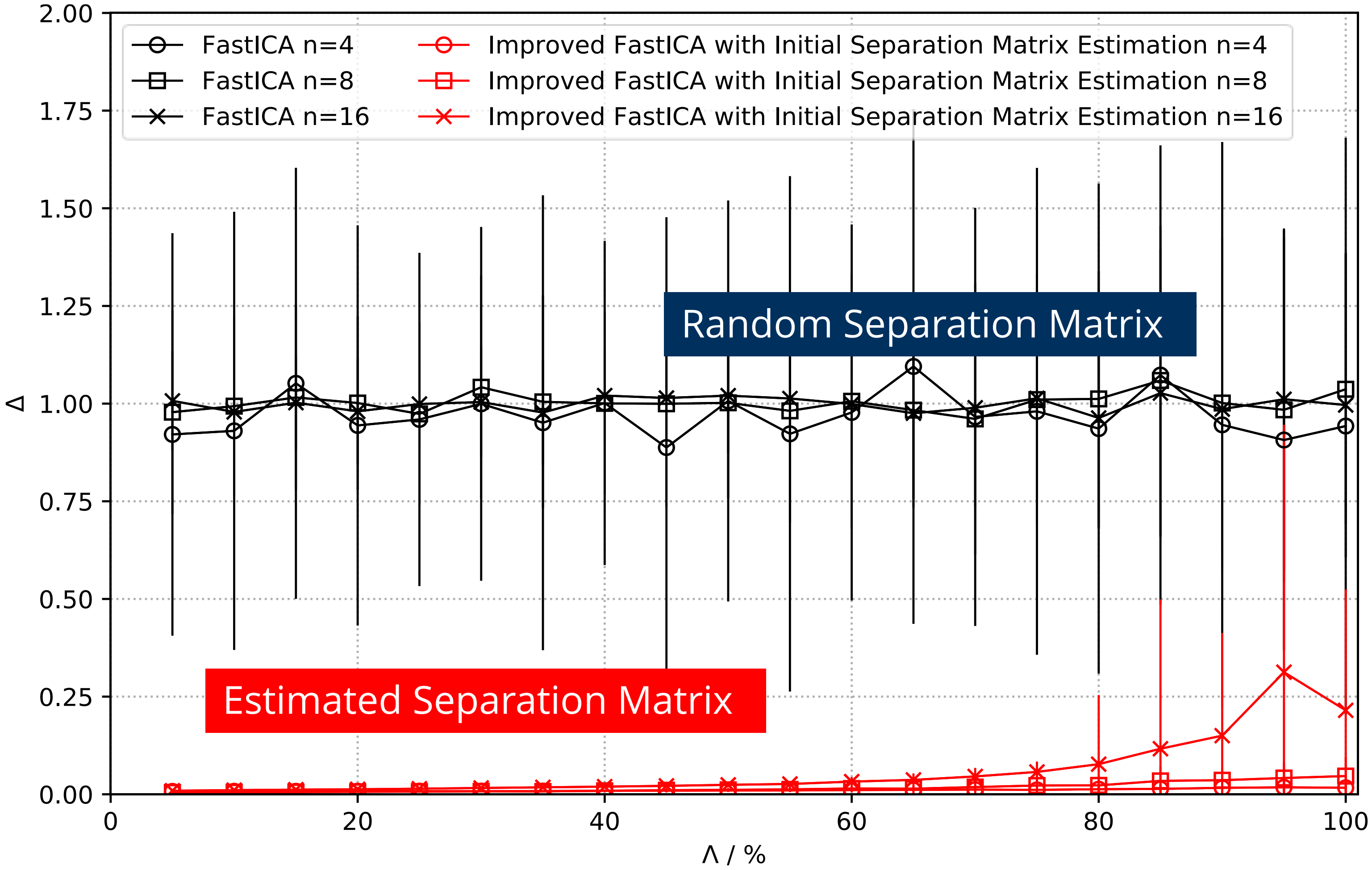
Sample size:

— 30 samples with 95% confidence level

Environment:

— INTEL-I5-6200U 2.8Ghz

4 Evaluation
Difference from Ideal Separation Matrix W



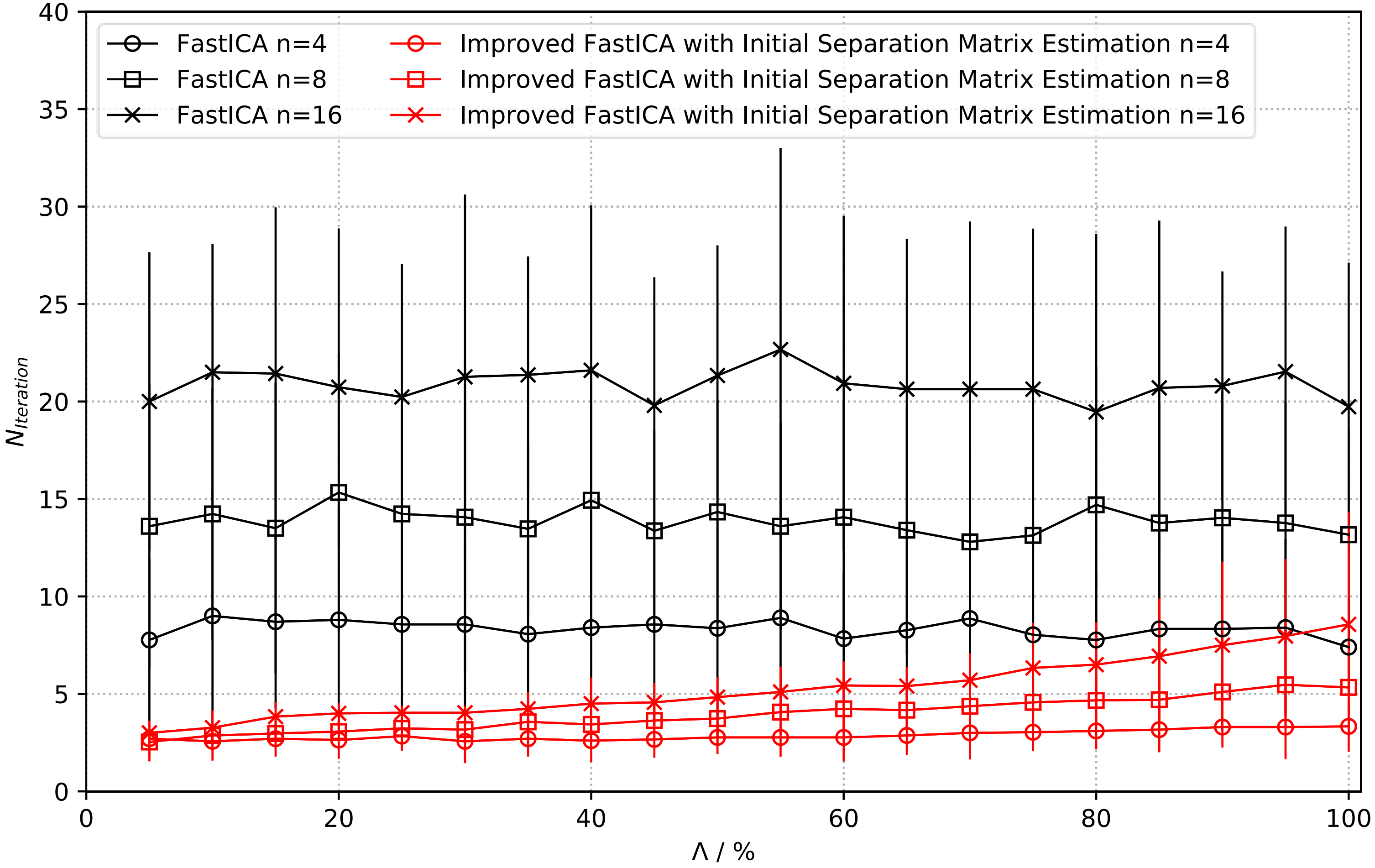
Cosine distance [9]:

$$\Delta = 1 - \frac{\boldsymbol{W}_{initial} \cdot \boldsymbol{W}}{\|\boldsymbol{W}_{initial}\|_2 \cdot \|\boldsymbol{W}\|_2}$$

- The estimated initial separation matrices have smaller differences from the ideal separation matrix

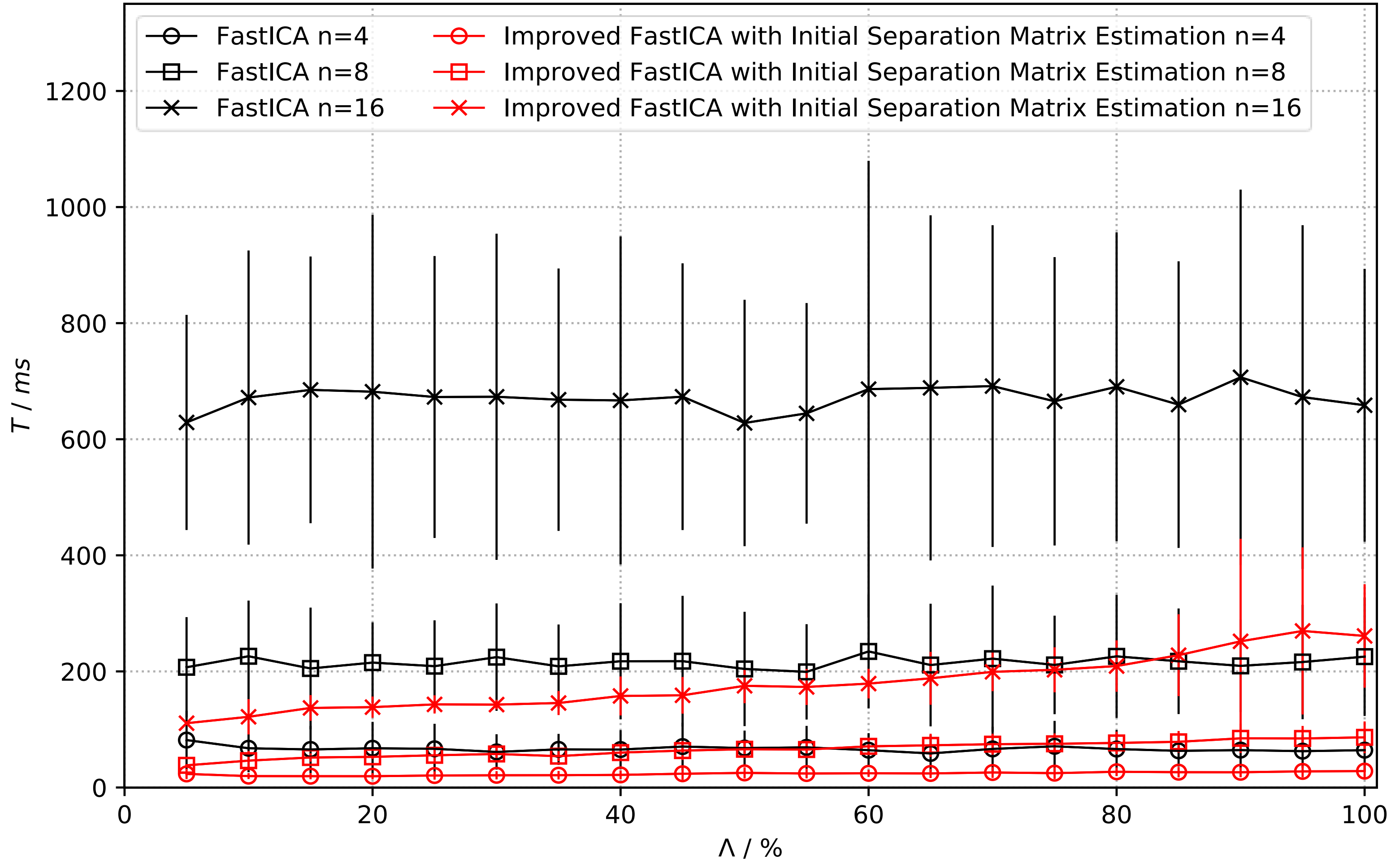
4 Evaluation

Iteration Number – Improved FastICA with Initial Separation Matrix Estimation



- The estimated initial matrices reduce the iteration numbers obviously
- The iteration number increases as Λ grows

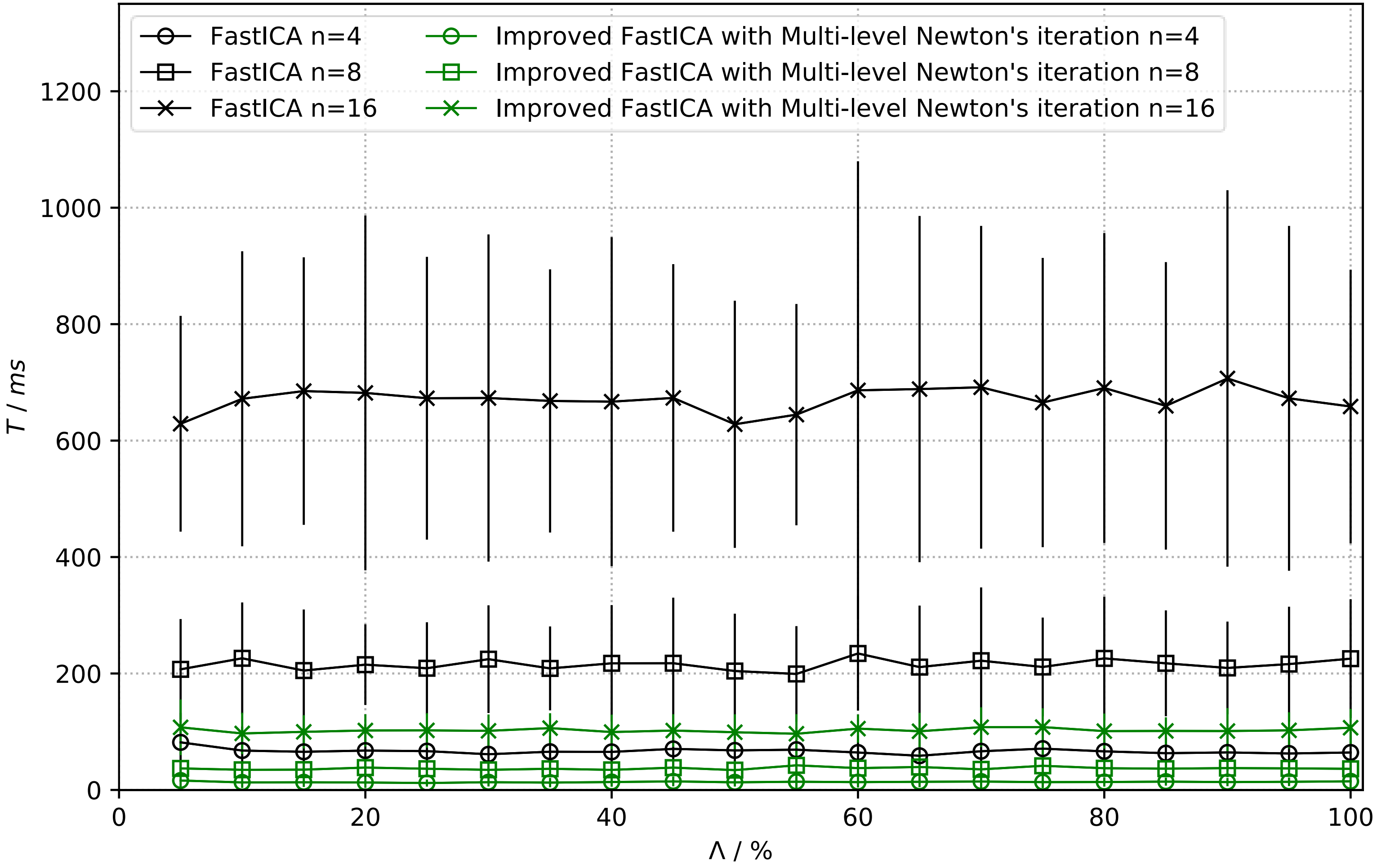
4 Evaluation
Time – Improved FastICA with Initial Separation Matrix Estimation



- The time consumption increases as Λ grows
- Better stability

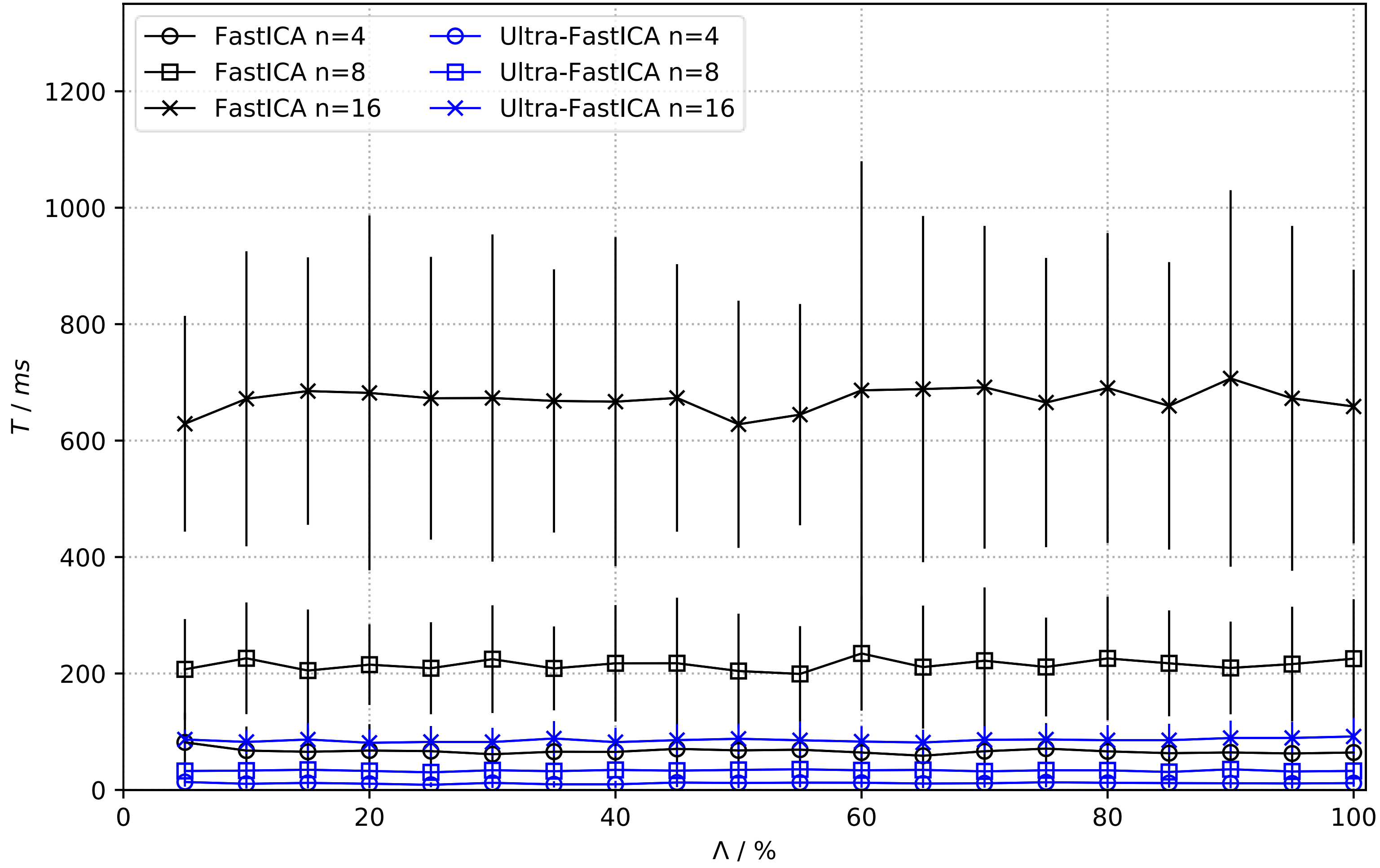
4 Evaluation

Time – Improved FastICA with Multi-Level Newton's Iteration



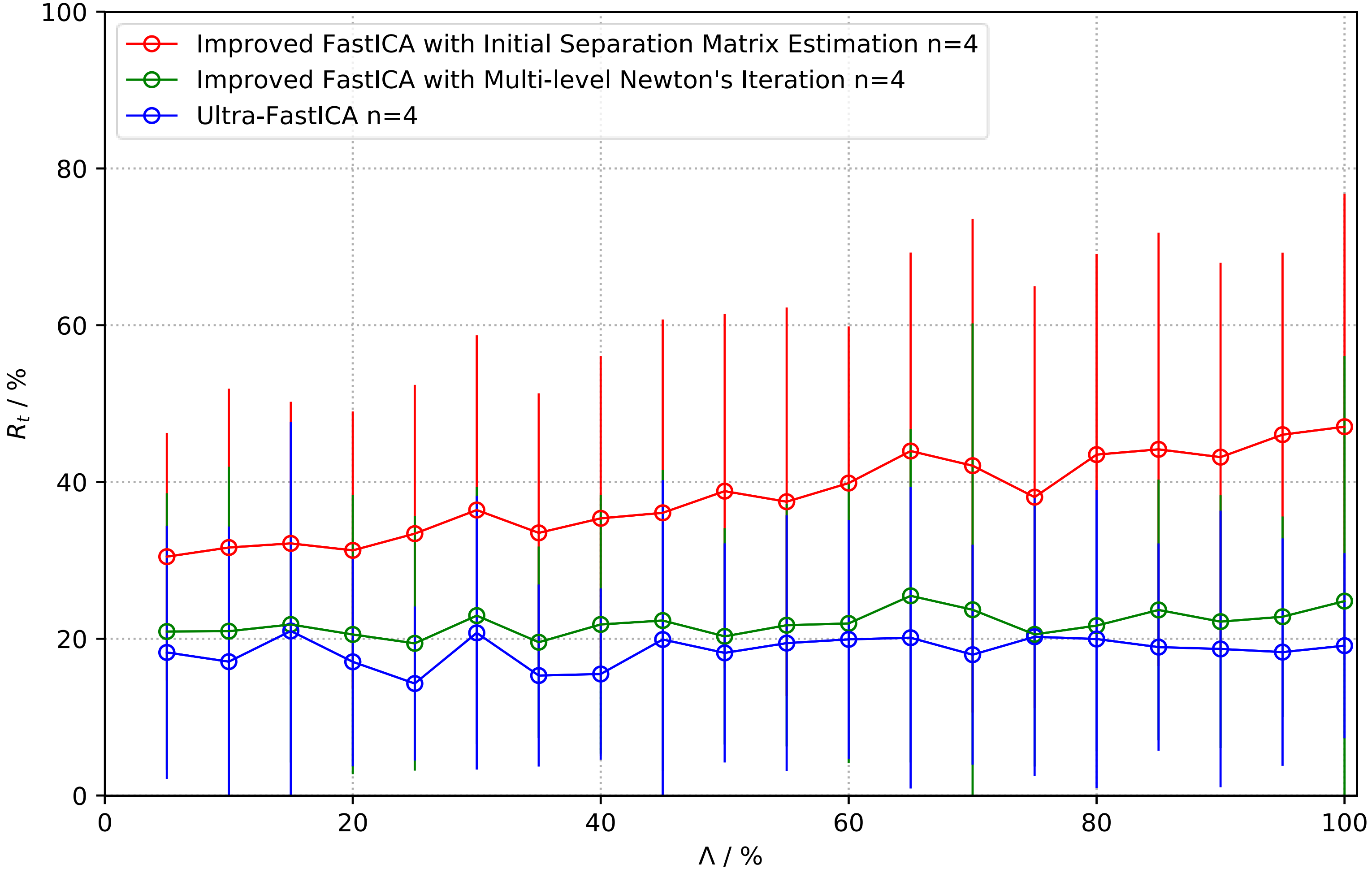
- The time consumptions are independent on Λ
- Better stability
- Faster

4 Evaluation
Time – Ultra-FastICA



- The time consumptions are independent on Λ
- Better stability
- Fastest

4 Evaluation
Time Consumption comparison, source number n=4

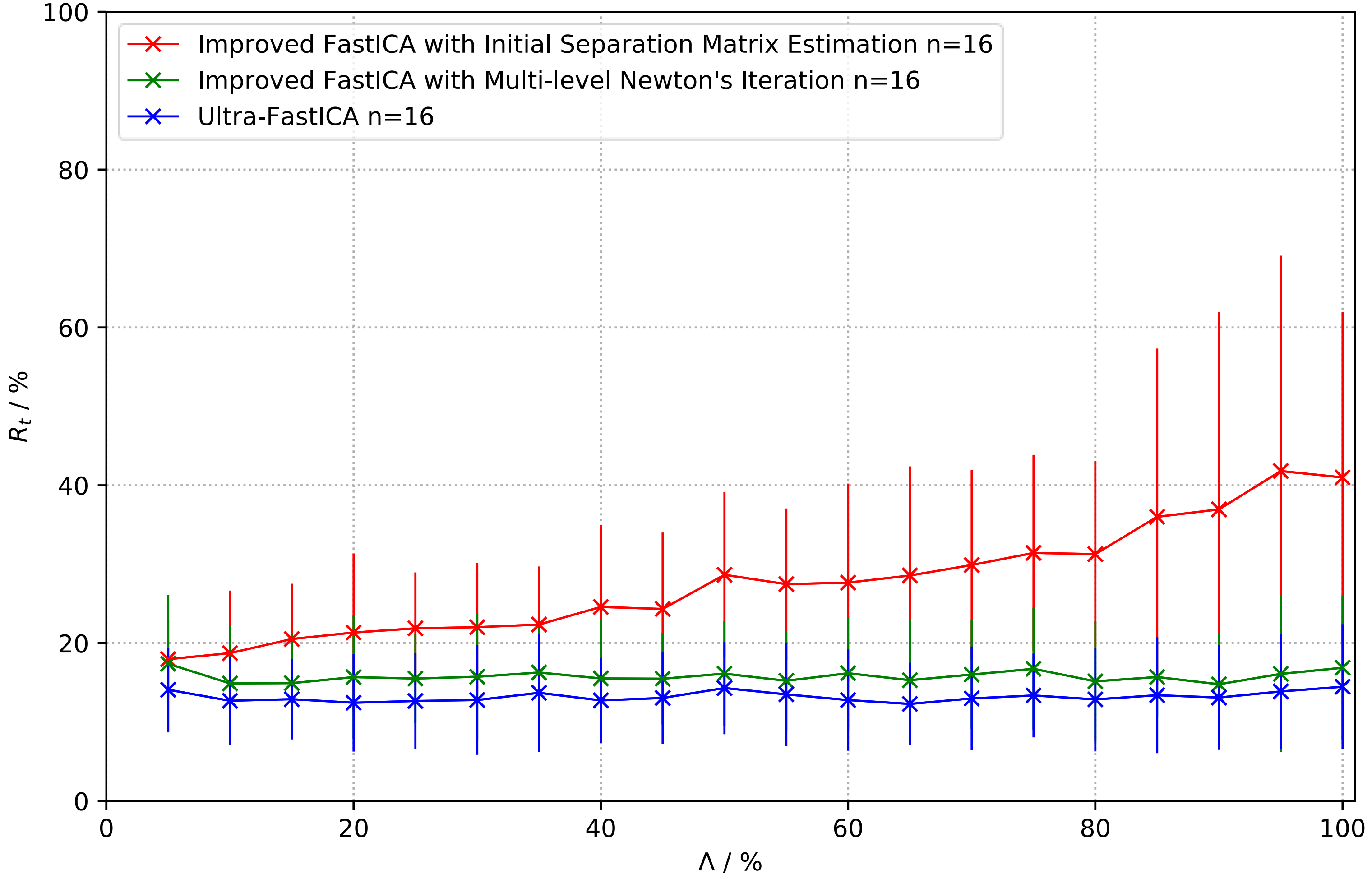


Ratio of time consumption:

$$R_t = \frac{T_{Tested\ BSS\ Algorithm}}{T_{FastICA}} \times 100\%$$

- Time consumption of Improved FastICA with Initial Separation Matrix Estimation increases as Λ grows
- Other two methods are more stable
- Ultra-FastICA has the best performance

4 Evaluation
Time Consumption comparison, source number n=16

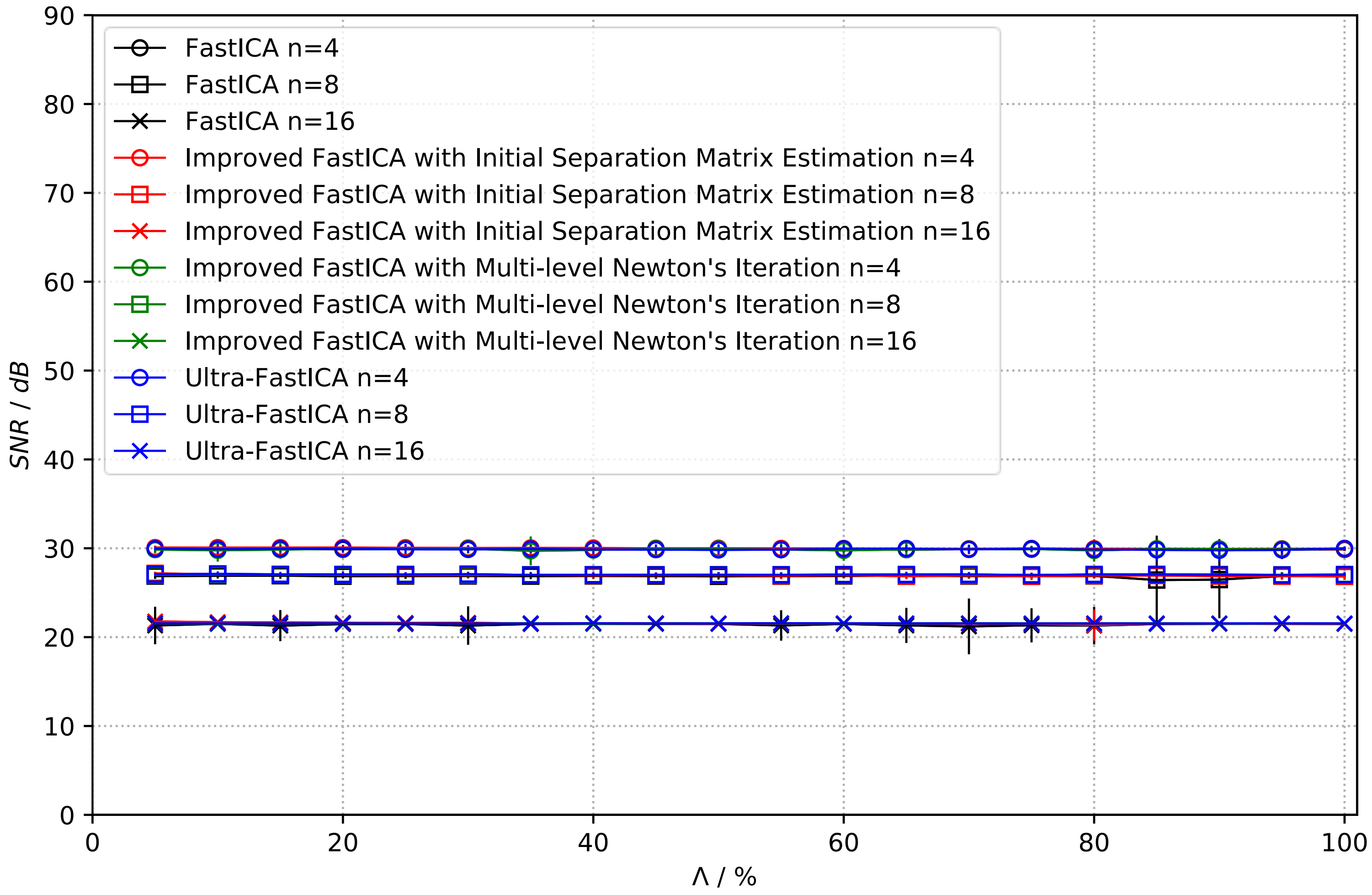


Ratio of time consumption:

$$R_t = \frac{T_{Tested\ BSS\ Algorithm}}{T_{FastICA}} \times 100\%$$

- Time consumption of Improved FastICA with Initial Separation Matrix Estimation increases as Λ grows
- Other two methods are more stable
- Ultra-FastICA has the best performance

4 Evaluation
SNR comparison



- The SNR are almost the same as FastICA

5 Conclusions
Fast Source Separation

- New directions for improving the FastICA:
  - — Iteration Number
  - — Computation

- Three novel ICA algorithms proposed:
  - — Improved FastICA with Initial Separation Matrix Estimation (18%-48%)
  - — Improved FastICA with Multi-Level Newton's Iteration (14%-26%)
  - — Ultra-FastICA (12%-21%)

- Can be integrated into other improved FastICA or ICA algorithm
- Low latency

6 References

[1] Handbook of Blind Source Separation: Independent component analysis and applications[M]. Academic press, 2010.

[2] Jung T P, Makeig S, McKeown M J, et al. Imaging brain dynamics using independent component analysis[J]. Proceedings of the IEEE, 2001, 89(7): 1107-1122.

[3] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications[J]. Neural networks, 2000, 13(4-5): 411-430.

[4] Ahmad T, Alias N, Ghanbari M, et al. Improved Fast ICA Algorithm Using Eighth-Order Newton's Method[J]. Research Journal of Applied Sciences, Engineering and Technology, 2013, 6(10): 1794-1798.

[5] Huang X. An improved FastICA algorithm for blind signal separation and its application[C]//2012 International Conference on Image Analysis and Signal Processing. IEEE, 2012: 1-4.

6 References

[6] Yuan L, Zhou Z, Yuan Y, et al. An improved FastICA method for fetal ECG extraction[J]. Computational and mathematical methods in medicine, 2018, 2018.

[7] Meng X, Yang L, Xie P, et al. An Improved FastICA Algorithm Based on Modified-M Estimate Function[J]. Circuits, Systems, and Signal Processing, 2018, 37(3): 1134-1144.

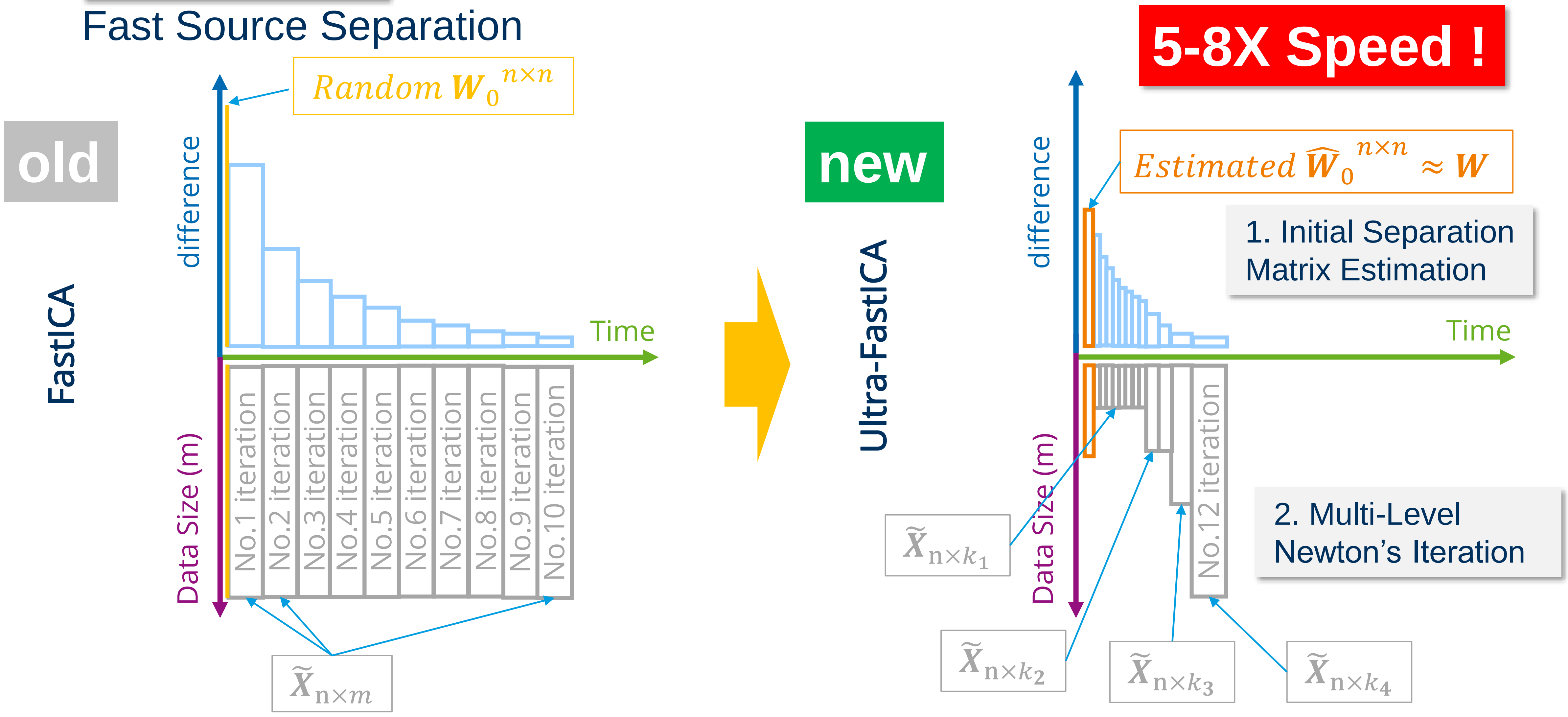
[8] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python[J]. Journal of machine learning research, 2011, 12(Oct): 2825-2830.

[9] Li B, Han L. Distance weighted cosine similarity measure for text classification[C]//International Conference on Intelligent Data Engineering and Automated Learning. Springer, Berlin, Heidelberg, 2013: 611-618.

Thank you!

Overview

Fast Source Separation



Appendix 1

Initial Separation Matrix Estimation

Precondition for Initial Separation Matrix Estimation:

— Each observer must close to different sources **Solution in Appendix 5**

For 2 sources, we have

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & a_{1,2} \\ a_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{bmatrix} \quad s.t.\, 0 < a_{1,2}, a_{2,1} < 1 \quad (12)$$

Let

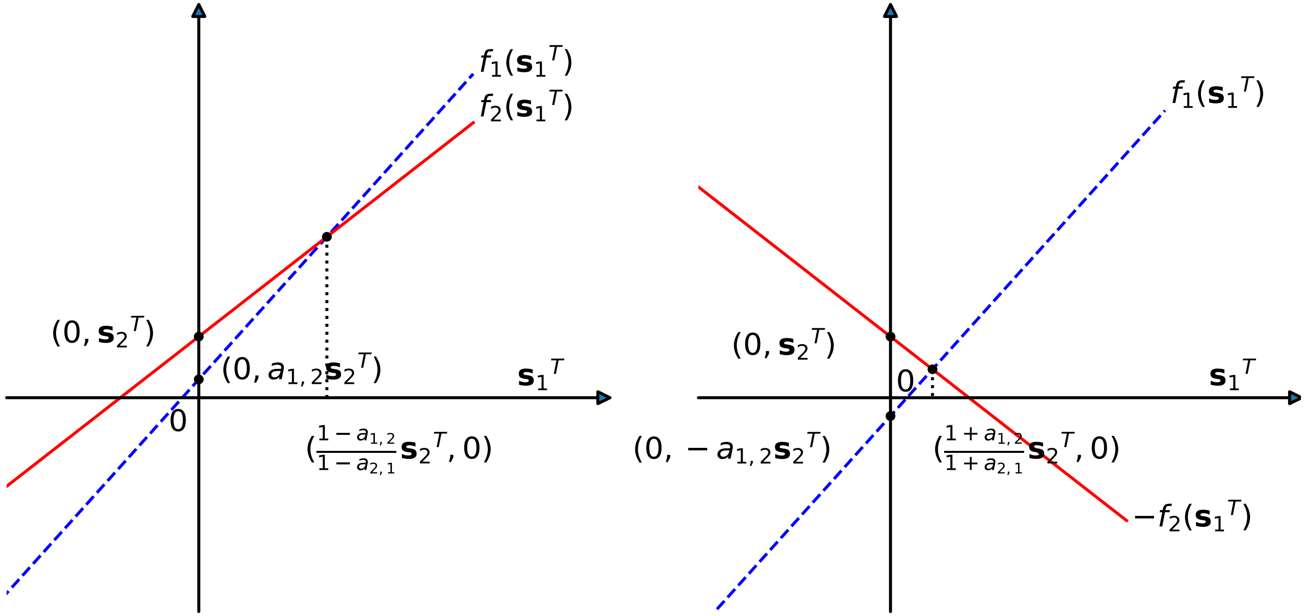
$$\begin{aligned} f_1(s_{1,j}) &= s_{1,j} + a_{1,2} s_{2,j} = x_{1,j} \\ f_2(s_{1,j}) &= a_{2,1} s_{2,j} + s_{2,j} = x_{2,j} \end{aligned} \quad (13)$$

When $f_1(s_{1,j}) > \pm f_2(s_{1,j})$, then we have

$$s_{1j} > \frac{1 - a_{1,2}}{1 - a_{2,1}} \geq 0 \quad s.t.\, s_{2,j} \geq 0 \quad (14)$$

$$s_{1,j} > -\frac{1 + a_{1,2}}{1 + a_{2,1}} \geq 0 \quad s.t.\, s_{2,j} < 0 \quad (15)$$

Initial Separation Matrix Estimation



Let

$$f_1(s_{1,j}) = s_{1,j} + a_{1,2}s_{2,j} = x_{1,j}$$
$$f_2(s_{1,j}) = a_{2,1}s_{2,j} + s_{2,j} = x_{2,j} \quad (13)$$

When $f_1(s_{1,j}) > \pm f_2(s_{1,j})$, then we have

$$s_{1j} > \frac{1-a_{1,2}}{1-a_{2,1}} \geq 0 \quad s.t. s_{2,j} \geq 0 \quad (14)$$

$$s_{1,j} > -\frac{1+a_{1,2}}{1+a_{2,1}} \geq 0 \quad s.t. s_{2,j} < 0 \quad (15)$$

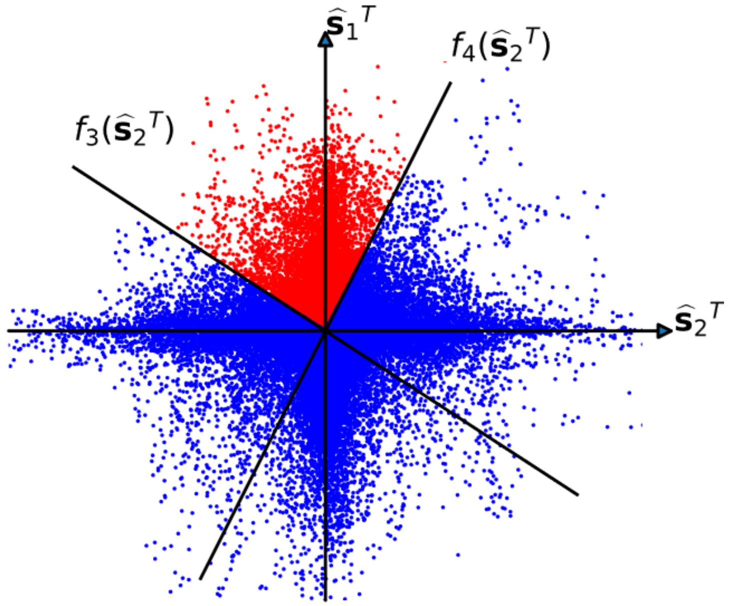
Initial Separation Matrix Estimation

When $\hat{\boldsymbol{X}}$ fulfills Eq.(18) and Eq.(19), we have:

$$\hat{\boldsymbol{X}} = \begin{bmatrix} x_{1,b_1} & \cdots & x_{1,b_t} \\ x_{2,b_1} & \cdots & x_{2,b_t} \end{bmatrix} \quad s.t. \quad \forall x_{1,b_i} > |x_{2,b_i}|, t < m \quad (16)$$

$$\begin{aligned} x_{1,b_i} + x_{2,b_i} &= f_1(s_{1,b_i}) + f_2(s_{1,b_i}) > 0 \\ x_{1,b_i} - x_{2,b_i} &= f_1(s_{1,b_i}) - f_2(s_{1,b_i}) > 0 \end{aligned} \quad (17)$$

$$\begin{cases} s_{1,b_i} > -\dfrac{1+a_{1,2}}{1+a_{2,1}} s_{2,b_i} = f_3(s_{2,b_i}) \geq 0 & s_{2,b_i} < 0 \\ s_{1,b_i} > \dfrac{1-a_{1,2}}{1-a_{2,1}} s_{2,b_i} = f_4(s_{2,b_i}) \geq 0 & s_{2,b_i} > 0 \end{cases} \quad (18)$$



$$\begin{cases} \exists s_{2,b_i} \geq 0 \\ \exists s_{2,b_i} \leq 0 \\ \forall s_{1,b_i} \geq 0 \end{cases} \quad (19)$$

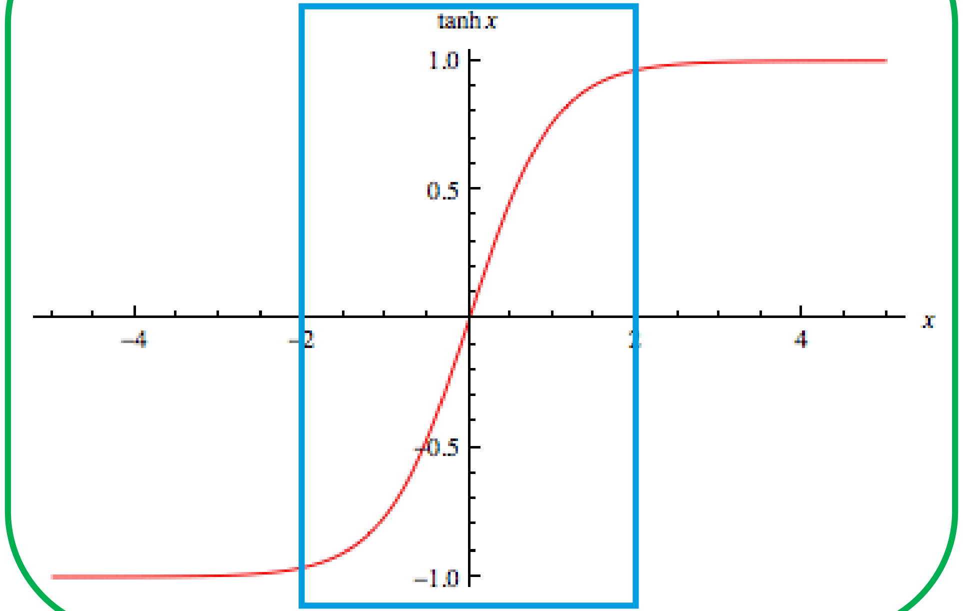
$$|E(s_{1,b_i})| \gg or > |E(s_{2,b_i})|$$

Initial Separation Matrix Estimation

The form of the FastICA algorithm in scikit-learn is as follows:

1. Centralization
2. Whiten, $\boldsymbol{\tilde{X}} = \boldsymbol{VX}$
3. Choose an initial (e.g. random) $\boldsymbol{W_0}$
4. Let $\boldsymbol{W}^{+} = \boldsymbol{E}\{\boldsymbol{\tilde{X}}g(\boldsymbol{W^T\tilde{X}})\} - \boldsymbol{E}\{g'(\boldsymbol{W^T\tilde{X}})\}\boldsymbol{W}$
5. Decorrelation, $\boldsymbol{W} = \boldsymbol{W}^{+}/||\boldsymbol{W}^{+}||$
6. If $Diff > Tol$, go back to 4, otherwise go to 7
7. Let $\boldsymbol{\hat{S}} = \boldsymbol{W\tilde{X}}$

$$g(\boldsymbol{W^T\tilde{X}}) = tanh(\boldsymbol{W^T\tilde{X}})$$



$\mathbf{Var}(\boldsymbol{W^T\tilde{X}}) \approx \mathbf{1}$

Appendix 2

Initial Separation Matrix Estimation

The form of the FastICA algorithm in scikit-learn is as follows:

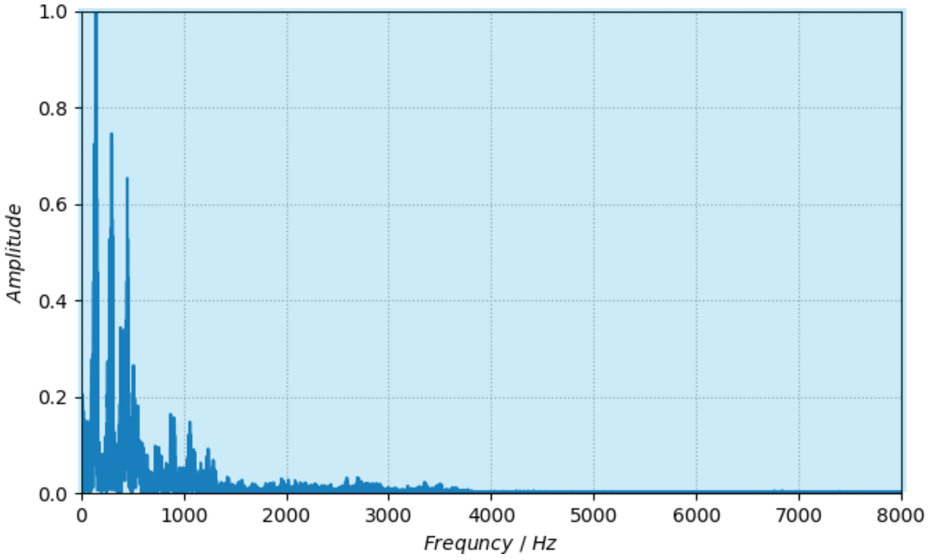
1. Centralization
2. Whiten, $\widetilde{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{X}$
3. Choose an initial (e.g. random) $\boldsymbol{W_0}$
4. Let $\boldsymbol{W}^+ = \boldsymbol{E}\{\widetilde{\boldsymbol{X}}g(\boldsymbol{W}^{\boldsymbol{T}}\widetilde{\boldsymbol{X}})\} - \boldsymbol{E}\{g'(\boldsymbol{W}^{\boldsymbol{T}}\widetilde{\boldsymbol{X}})\}\boldsymbol{W}$
5. Decorrelation, $\boldsymbol{W} = \boldsymbol{W}^+/\|\boldsymbol{W}^+\|$
6. If $Diff > Tol$, go back to 4, otherwise go to 7
7. Let $\widehat{\boldsymbol{S}} = \boldsymbol{W}\widetilde{\boldsymbol{X}}$

The norm of W must be the same

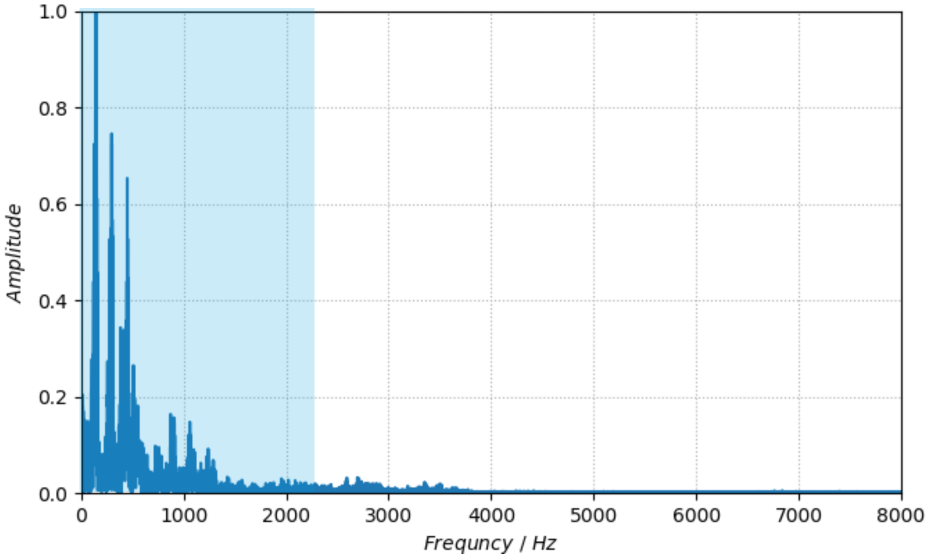
$$Estimated\ intial\ \widehat{\boldsymbol{W}}_0{}^{n\times n} = \frac{\widehat{\boldsymbol{W}}}{\|\widehat{\boldsymbol{W}}\|}$$

Multi-Level Newton’s Iteration

Frequency response of the source signal fs=16000Hz:

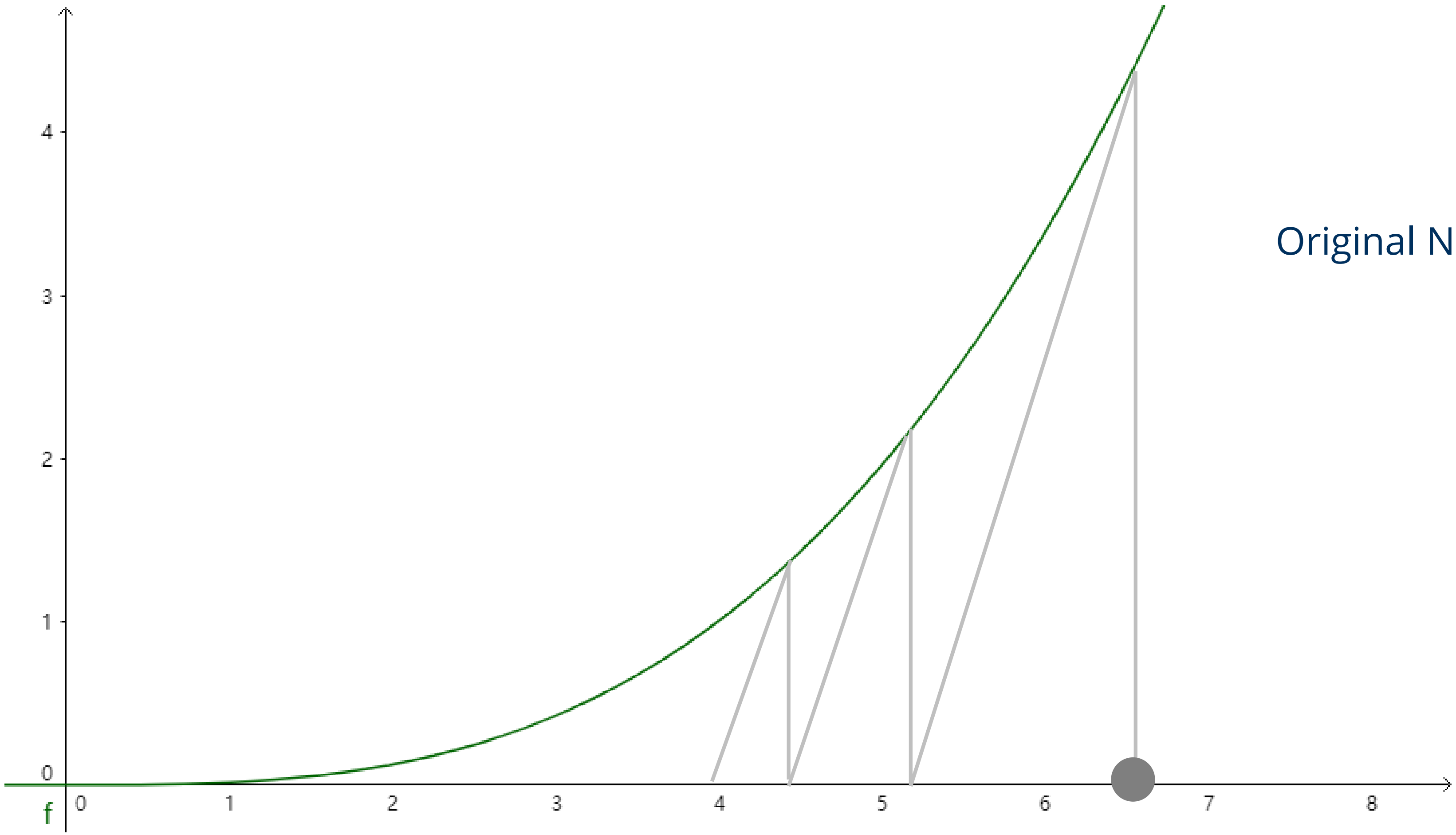


Original signal



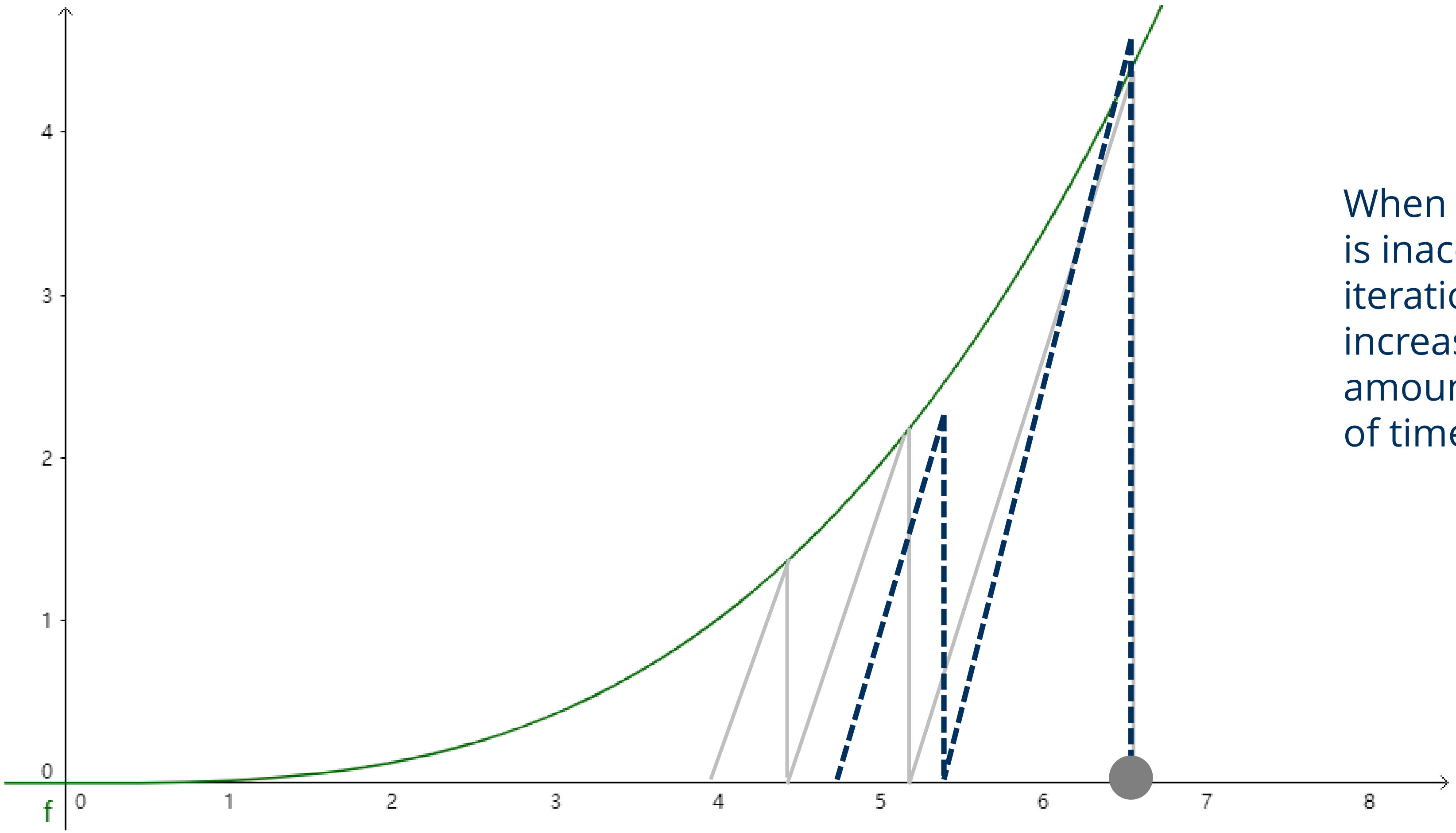
Partial signal

Improvement for Newton's Iteration



Original Newton's Iteration

Improvement for Newton's Iteration

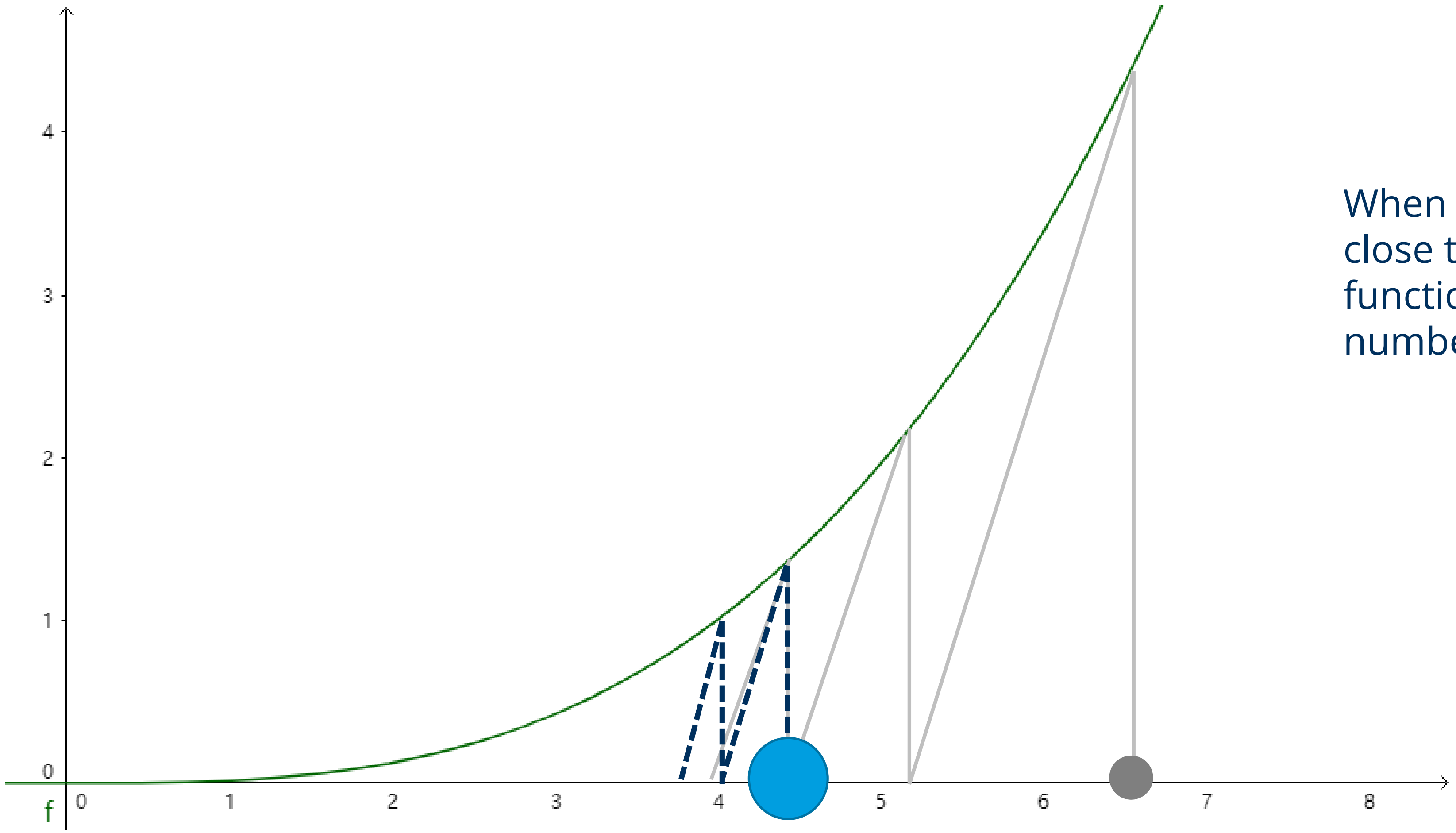


When the computation is inaccurate, the iteration number will increase by a small amount but save a lot of time of computation

Appendix 3

Improvement for Newton's Iteration

2. Multi-Level Newton's Iteration



When the start point is close to the root of the function, the iteration number will decrease

Multi-Level Newton's Iteration

Assume that $\boldsymbol{W}_{\widetilde{\boldsymbol{X}}'}$ is the separation matrix computed by whitened mixed signal $\widetilde{\boldsymbol{X}}$, then

$$\boldsymbol{W}_{\widetilde{\boldsymbol{X}}} = \widehat{\boldsymbol{S}}\widetilde{\boldsymbol{X}}'^{-1} = \widehat{\boldsymbol{S}}\widetilde{\boldsymbol{X}}^{T} \quad (20)$$

So that the value of $\boldsymbol{W}_{{}_{e}^{i}\widetilde{\boldsymbol{X}}}$ is

$$\boldsymbol{W}_{{}_{e}^{i}\widetilde{\boldsymbol{X}}} = e\,{}_{e}^{i}\widehat{\boldsymbol{S}}\,{}_{e}^{i}\widetilde{\boldsymbol{X}}^{T} \quad (21)$$

Together with the above equations, we have

$$\boldsymbol{W}_{\widetilde{\boldsymbol{X}}} = \widehat{\boldsymbol{S}}\widetilde{\boldsymbol{X}}^{-1} = \frac{1}{e}\sum_{i=1}^{e} e\,{}_{e}^{i}\widehat{\boldsymbol{S}}\,{}_{e}^{i}\widetilde{\boldsymbol{X}}^{T} = \frac{1}{e}\sum_{i=1}^{e} \boldsymbol{W}_{{}_{e}^{i}\widetilde{\boldsymbol{X}}} \quad (22)$$

So that

$$\sigma_{\boldsymbol{W}_{\widetilde{\boldsymbol{X}}}} = \frac{1}{\sqrt{e}}\sigma_{\boldsymbol{W}_{{}_{e}^{i}\widetilde{\boldsymbol{X}}}} \quad (23)$$

Multi-Level Newton's Iteration

To determine the convergence tolerance of Newton's iteration, let

$$\boldsymbol{W}^{+} = \boldsymbol{W} = \boldsymbol{W}_{\boldsymbol{ideal}} + \sigma \quad (24)$$

Then we have

$$\boldsymbol{W}^{+}\boldsymbol{W} - \boldsymbol{I} = (\boldsymbol{W}_{\boldsymbol{ideal}} + \sigma)^2 - \boldsymbol{I} = 2\sigma\boldsymbol{W}_{\boldsymbol{ideal}} + \sigma^{\boldsymbol{2}} \approx 2\sigma\boldsymbol{W}_{\boldsymbol{ideal}} \propto \sigma \propto Diff \quad (25)$$

According to $\sigma_{\boldsymbol{W}_{\widetilde{\boldsymbol{X}'}}} = \frac{1}{\sqrt{e}}\sigma_{\boldsymbol{W}_{{}_{e}^{i}\widetilde{\boldsymbol{X}'}}}$, then the tolerance of difference for extracted signal ${}_{e}^{i}\widetilde{\boldsymbol{X}'}$ is

$$Tol_{\widetilde{\boldsymbol{X}}} \approx \frac{1}{\sqrt{e}} Tol_{{}_{e}^{i}\widetilde{\boldsymbol{X}}} \quad \rightarrow \quad Tol_{{}_{e}^{i}\widetilde{\boldsymbol{X}}} \approx \sqrt{e} \cdot Tol_{\widetilde{\boldsymbol{X}}} \quad (26)$$

Improvement

Improvement for Initial Separation Matrix Estimation

— A constant separation matrix $\boldsymbol{W'}$ is used for roughly source separation

$$\widehat{\boldsymbol{A}}' = \begin{bmatrix} 1 & \cdots & \hat{a}_{1,n} \\ \hat{a}_{2,1} & \cdots & \hat{a}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{a}_{n,1} & \cdots & 1 \end{bmatrix} = \boldsymbol{w}'^{-1}\boldsymbol{AK}$$

Signal Extraction Interval Determination

Extraction interval for signal extraction is according to the $Diff$ and $Tol_{{}_{e}^{i}\widetilde{\mathbf{X}}}$, the $Diff$ is calculated by Newton's iteration.

The extraction interval e is the maximum value of e under the condition $Diff > Tol_{{}_{e}^{i}\widetilde{\mathbf{X}}}$. It is no need to get a very small tolerance with a big e, which will cost a lot of time and is useless.

However, when we according to principle above to select the extraction interval, the separation quality is not so optimal, so that the calculated $Diff$ will be corrected to let the value of $Diff$ more reliable, $Diff$ between two extraction interval can be calculated as

$$Diff_{{}_{e}^{i}\widetilde{\mathbf{X}}} \equiv \frac{\sqrt{e'}}{\sqrt{e}} Diff_{{}_{e'}^{i}\widetilde{\mathbf{X}}}$$

Multi-Level Newton's Iteration – Algorithm

1. Centralization & Whiten, $\widetilde{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{X}$
2. Choose an initial (e.g. random) $\widehat{\boldsymbol{W}}_0$ and $\boldsymbol{W} = \widehat{\boldsymbol{W}}_0$
3. Choose an initial extraction interval $\mathrm{e} = e_0$
4. Let $\boldsymbol{W}^{+}, Diff = f_{Newton's\ iteration}\left(\widetilde{\tfrac{1}{e}\boldsymbol{X}}, \boldsymbol{W}, \sqrt{e} \cdot Tol\right)$
5. Let $Diff' = Diff \times \sqrt{e}$
6. Let $\mathrm{e} = \mathrm{e} - 1$
7. If $\mathrm{e} > 0$ and $\frac{Diff'}{\sqrt{e}} \geq \sqrt{e} \cdot Tol$, go back to 4, if $\mathrm{e} > 0$ and $\frac{Diff'}{\sqrt{e}} < \sqrt{e} \cdot Tol$, go back to 6, otherwise go to 8
8. Get the separation matrix $\boldsymbol{W}$ and let $\widehat{\boldsymbol{S}} = \boldsymbol{W}\widetilde{\boldsymbol{X}}$

Convergence Tolerance

Mixed Signals

Initial W

FastICA

Multi-Level Newton's Iteration

The performance of Improved FastICA with Multi-Level Newton's Iteration with different initial signal extraction interval e_0, n=16