

VARIABLE PROJECTION FOR MULTIPLE FREQUENCY ESTIMATION

Yuneisy E. Garcia Guzman, Péter Kovács, Mario Huemer

Johannes Kepler University Linz, Institute of Signal Processing, Austria
Eötvös Loránd University, Department of Numerical Analysis, Hungary

May, 2020

Introduction

- Frequency estimation is a crucial field in many areas of signal processing such as audio signal processing, and radar. It is also related to direction of arrival (DoA) estimation.

Introduction

- Frequency estimation is a crucial field in many areas of signal processing such as audio signal processing, and radar. It is also related to direction of arrival (DoA) estimation.
- Several algorithms have been proposed in the literature: MUSIC, Root-MUSIC, ESPRIT → **high computational complexity**.

R. Schmidt. Multiple emitter location and signal parameter estimation. IEEE Transactions Antennas and Propagation, 2011.

B.D Rao, K. Hari. Performance analysis of Root-MUSIC. IEEE Transactions on Acoustics, Speech and Signal Processing, 1989.

A. Paulraj, R. Roy, T. Kailath. Estimation of signal parameter via rotational invariance techniques-ESPRIT. Circuits Systems and Computers, 1985.

Introduction

- Frequency estimation is a crucial field in many areas of signal processing such as audio signal processing, and radar. It is also related to direction of arrival (DoA) estimation.
- Several algorithms have been proposed in the literature: MUSIC, Root-MUSIC, ESPRIT → high computational complexity.
- The discrete-time Fourier transform (DTFT) → low complexity, but becomes biased due to the interactions of the different frequencies.

R. Schmidt. Multiple emitter location and signal parameter estimation. IEEE Transactions Antennas and Propagation, 2011.

B.D Rao, K. Hari. Performance analysis of Root-MUSIC. IEEE Transactions on Acoustics, Speech and Signal Processing, 1989.

A. Paulraj, R. Roy, T. Kailath. Estimation of signal parameter via rotational invariance techniques-ESPRIT. Circuits Systems and Computers, 1985.

Introduction

- Golub and Pereyra stated that frequency estimation can be formulated as a separable nonlinear least squares (SNLLS) fitting problems.

Introduction

- Golub and Pereyra stated that frequency estimation can be formulated as a separable nonlinear least squares (SNLLS) fitting problems.
- Such representation encourages the use of VP-like algorithms.

Introduction

- Golub and Pereyra stated that frequency estimation can be formulated as a separable nonlinear least squares (SNLLS) fitting problems.
- Such representation encourages the use of VP-like algorithms.
- Singular value decomposition (SVD) methods → high computational complexity.

Introduction

- We provide a formulation of the frequency estimation problem in the framework of Γ -approximation → quantify the difficulty of the optimization.

Introduction

- We provide a formulation of the frequency estimation problem in the framework of Γ -approximation → quantify the difficulty of the optimization.
- We propose an alternative method that accelerates the computation of the exact gradient.

Introduction

- We provide a formulation of the frequency estimation problem in the framework of Γ -approximation → quantify the difficulty of the optimization.
- We propose an alternative method that accelerates the computation of the exact gradient.
- Comparisons with other state-of-the-art methods are also provided.

Numerical analysis

Let's consider the γ -polynomials of order n :

$$\sigma(\mathbf{c}, \mathbf{f}) \equiv \sigma(\mathbf{c}, \mathbf{f}; t) = \sum_{j=1}^n c_j \gamma(f_j; t) \quad (t \in \mathbb{R}), \quad (1)$$

- $\mathbf{c} \in \mathbb{C}^n$.
- $a < f_1 < f_2 < \dots < f_n < b$ is a subdivision of the interval (a, b) by n distinct points.

Numerical analysis

$$\sigma(\mathbf{c}, \mathbf{f}) \equiv \sigma(\mathbf{c}, \mathbf{f}; t) = \sum_{j=1}^n c_j \gamma(f_j; t) \quad (t \in \mathbb{R}) \quad (1)$$

- $\sigma(\mathbf{c}, \mathbf{f}; t)$ is the nonlinear model of a complex valued signal y .
- the coefficients c_j 's denote complex amplitudes.
- $f_j \in [0, 0.5]$ are the frequencies.
- $\gamma(f_j; t) = \exp(i2\pi f_j t)$.

VP gradient

$$\arg \min_{(\mathbf{c}, \mathbf{f}) \in \mathbb{C}^n \times s_n} F(\mathbf{c}, \mathbf{f}) = \arg \min_{(\mathbf{c}, \mathbf{f}) \in \mathbb{C}^n \times s_n} \|y - \sigma(\mathbf{c}, \mathbf{f})\|_2^2 \quad (2)$$

- $\mathbf{c}(\mathbf{f}) = \Psi^\dagger(\mathbf{f})\mathbf{y}$ is the minimal least squares solution for a fixed \mathbf{f} .
- $\Psi^\dagger(\mathbf{f})$ is the Moore–Penrose pseudoinverse.
- $\Psi(\mathbf{f}) = [\Psi_1, \dots, \Psi_n]$ denotes the matrix functions and $\Psi_1 = [1, e^{i2\pi f_1}, \dots, e^{i2\pi f_1(M-1)}]$.
- $\mathbf{P}_{\Psi(\mathbf{f})} = \Psi(\mathbf{f})\Psi^\dagger(\mathbf{f})$ is the orthogonal projector on the linear space spanned by the columns of the matrix $\Psi(\mathbf{f})$.
- $\mathbf{P}_{\Psi(\mathbf{f})}^\perp = \mathbf{I} - \mathbf{P}_{\Psi(\mathbf{f})}$ denotes the projector on the orthogonal complement of the column space of $\Psi(\mathbf{f})$.

VP gradient

- The full functional problem:

$$\arg \min_{(\mathbf{c}, \mathbf{f}) \in \mathbb{C}^n \times s_n} \tilde{F}(\mathbf{c}, \mathbf{f}) = \arg \min_{(\mathbf{c}, \mathbf{f}) \in \mathbb{C}^n \times s_n} \|y - \sigma(\mathbf{c}, \mathbf{f})\|_2^2 \quad (2)$$

- $\mathbf{c}(\mathbf{f}) = \Psi^\dagger(\mathbf{f})\mathbf{y}$ is the minimal least squares solution for a fixed \mathbf{f} .

- The reduced functional problem:

$$\arg \min_{\mathbf{f} \in s_n} \tilde{F}(\mathbf{f}) = \arg \min_{\mathbf{f} \in s_n} \|y - \sigma(\mathbf{c}(\mathbf{f}), \mathbf{f})\|_2^2 \quad (3)$$

VP gradient

Then the frequency parameters can be calculated by solving the following optimization

$$\arg \min_{\mathbf{f} \in s_n} \left\| \mathbf{y} - \Psi(\mathbf{f}) \Psi^\dagger(\mathbf{f}) \mathbf{y} \right\|_2^2 = \arg \min_{\mathbf{f} \in s_n} \left\| \mathbf{P}_{\Psi(\mathbf{f})}^\perp \mathbf{y} \right\|_2^2 \quad (4)$$

The resulting functional is a VP functional.

G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413–432, 1973.

D. P. O'Leary and B. W. Rust, "Variable Projection for Nonlinear Least Squares Problems," *Computational Optimization and Applications*, vol. 54, no. 3, pp. 579–593, 2013.

VP gradient

Then the frequency parameters can be calculated by solving the following optimization

$$\arg \min_{\mathbf{f} \in s_n} \|\mathbf{y} - \Psi(\mathbf{f})\Psi^\dagger(\mathbf{f})\mathbf{y}\|_2^2 = \arg \min_{\mathbf{f} \in s_n} \left\| \mathbf{P}_{\Psi(\mathbf{f})}^\perp \mathbf{y} \right\|_2^2 \quad (4)$$

The resulting functional is a VP functional.

The k th coordinate of the gradient of the functional is given by

$$\frac{1}{2} \nabla \tilde{F}_k = \left[-(\mathbf{P}_\Psi \mathbf{D}_k \Psi^\dagger + (\mathbf{P}_\Psi \mathbf{D}_k \Psi^\dagger)^T) \mathbf{y} \right]^T \mathbf{P}_{\Psi}^\perp \mathbf{y}, \quad (5)$$

where $\mathbf{D}_k = \partial \Psi(\mathbf{f}) / \partial f_k$ represents the matrix of partial derivatives of $\Psi(\mathbf{f})$ with respect to the single parameter f_k .

G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 413–432, 1973.

D. P. O'Leary and B. W. Rust, "Variable Projection for Nonlinear Least Squares Problems," *Computational Optimization and Applications*, vol. 54, no. 3, pp. 579–593, 2013.

- Real representation of the problem (splitting into real $\mathcal{R}(\cdot)$ and imaginary $\mathcal{I}(\cdot)$ components):

$$\tilde{\Psi} = \begin{bmatrix} \mathcal{R}(\Psi) & -\mathcal{I}(\Psi) \\ \mathcal{I}(\Psi) & \mathcal{R}(\Psi) \end{bmatrix} \quad \tilde{\mathbf{y}} = \begin{bmatrix} \mathcal{R}(\mathbf{y}) \\ \mathcal{I}(\mathbf{y}) \end{bmatrix} \quad \tilde{\mathbf{c}} = \begin{bmatrix} \mathcal{R}(\mathbf{c}) \\ \mathcal{I}(\mathbf{c}) \end{bmatrix}.$$

- The SVD is replaced with a faster iterative calculation of the pseudoinverse.

$$\tilde{\Psi}^\dagger = (\tilde{\Psi}^T \tilde{\Psi})^{-1} \tilde{\Psi}^T$$

$$\mathbf{W}_Q = \frac{2}{M} \tilde{\Psi}^T \tilde{\Psi} \rightarrow \underbrace{\mathbf{W}_Q^{-1}}_{\text{MATRIX INVERSION LEMMA}}$$

$$\tilde{\Psi}^\dagger = \mathbf{W}_Q^{-1} \tilde{\Psi}^T$$

Iterative calculation of the pseudoinverse (ICP)

Algorithm 1 Iterative Computation of Pseudoinverse

Input: $\tilde{\Psi}$

Output: $\tilde{\Psi}^\dagger$

- 1: $[M, Q] = \text{size}(\tilde{\Psi})$
 - 2: $\mathbf{W}_Q = \frac{2}{M} \tilde{\Psi}^T \tilde{\Psi}$
 - 3: $\mathbf{W}_{Q-1}^{-1} \leftarrow \mathbf{1}$
 - 4: **for** $k = 2 : Q$ **do**
 - 5: $\mathbf{x}_Q \leftarrow \mathbf{W}_Q(1 : k - 1, k)$
 - 6: $\delta_Q \leftarrow \mathbf{x}_Q^T \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$
 - 7: $\mathbf{0} \leftarrow \text{zeros}(\text{length}(\mathbf{x}_Q), 1)$
 - 8: $\mathbf{y}_Q \leftarrow \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$
 - 9: $\mathbf{W}_{Q-1}^{-1} \leftarrow \begin{bmatrix} \mathbf{W}_{Q-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \frac{1}{1-\delta_Q} \begin{bmatrix} \mathbf{y}_Q \mathbf{y}_Q^T & -\mathbf{y}_Q \\ -\mathbf{y}_Q^H & 1 \end{bmatrix}$
 - 10: **end for**
 - 11: $\tilde{\Psi}^\dagger = \frac{1}{2M} \mathbf{W}_{Q-1}^{-1} \tilde{\Psi}^T$
-

Iterative calculation of the pseudoinverse (ICP)

Algorithm 1 Iterative Computation of Pseudoinverse

Input: $\tilde{\Psi}$

Output: $\tilde{\Psi}^\dagger$

```

1:  $[M, Q] = \text{size}(\tilde{\Psi})$ 
2:  $\mathbf{W}_Q = \frac{2}{M} \tilde{\Psi}^T \tilde{\Psi}$ 
3:  $\mathbf{W}_{Q-1}^{-1} \leftarrow \mathbf{1}$ 
4: for  $k = 2 : Q$  do
5:    $\mathbf{x}_Q \leftarrow \mathbf{W}_Q(1 : k - 1, k)$ 
6:    $\delta_Q \leftarrow \mathbf{x}_Q^T \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$ 
7:    $\mathbf{0} \leftarrow \text{zeros}(\text{length}(\mathbf{x}_Q), 1)$ 
8:    $\mathbf{y}_Q \leftarrow \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$ 
9:    $\mathbf{W}_{Q-1}^{-1} \leftarrow \begin{bmatrix} \mathbf{W}_{Q-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \frac{1}{1 - \delta_Q} \begin{bmatrix} \mathbf{y}_Q \mathbf{y}_Q^T & -\mathbf{y}_Q \\ -\mathbf{y}_Q^H & 1 \end{bmatrix}$ 
10: end for
11:  $\tilde{\Psi}^\dagger = \frac{1}{2M} \mathbf{W}_{Q-1}^{-1} \tilde{\Psi}^T$ 

```

Initialization

Iterative calculation of the pseudoinverse (ICP)

Algorithm 1 Iterative Computation of Pseudoinverse

Input: $\tilde{\Psi}$

Output: $\tilde{\Psi}^\dagger$

1: $[M, Q] = \text{size}(\tilde{\Psi})$

2: $\mathbf{W}_Q = \frac{2}{M} \tilde{\Psi}^T \tilde{\Psi}$

3: $\mathbf{W}_{Q-1}^{-1} \leftarrow \mathbf{1}$

4: **for** $k = 2 : Q$ **do**

5: $\mathbf{x}_Q \leftarrow \mathbf{W}_Q(1 : k - 1, k)$

6: $\delta_Q \leftarrow \mathbf{x}_Q^T \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$

7: $\mathbf{0} \leftarrow \text{zeros}(\text{length}(\mathbf{x}_Q), 1)$

8: $\mathbf{y}_Q \leftarrow \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$

9: $\mathbf{W}_{Q-1}^{-1} \leftarrow \begin{bmatrix} \mathbf{W}_{Q-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \frac{1}{1 - \delta_Q} \begin{bmatrix} \mathbf{y}_Q \mathbf{y}_Q^T & -\mathbf{y}_Q \\ -\mathbf{y}_Q^H & 1 \end{bmatrix}$

10: **end for**

11: $\tilde{\Psi}^\dagger = \frac{1}{2M} \mathbf{W}_{Q-1}^{-1} \tilde{\Psi}^T$

Matrix Inversion Lemma

Iterative calculation of the pseudoinverse (ICP)

Algorithm 1 Iterative Computation of Pseudoinverse

Input: $\tilde{\Psi}$

Output: $\tilde{\Psi}^\dagger$

- 1: $[M, Q] = \text{size}(\tilde{\Psi})$
- 2: $\mathbf{W}_Q = \frac{2}{M} \tilde{\Psi}^T \tilde{\Psi}$
- 3: $\mathbf{W}_{Q-1}^{-1} \leftarrow 1$
- 4: **for** $k = 2 : Q$ **do**
- 5: $\mathbf{x}_Q \leftarrow \mathbf{W}_Q(1 : k - 1, k)$
- 6: $\delta_Q \leftarrow \mathbf{x}_Q^T \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$
- 7: $\mathbf{0} \leftarrow \text{zeros}(\text{length}(\mathbf{x}_Q), 1)$
- 8: $\mathbf{y}_Q \leftarrow \mathbf{W}_{Q-1}^{-1} \mathbf{x}_Q$
- 9: $\mathbf{W}_{Q-1}^{-1} \leftarrow \begin{bmatrix} \mathbf{W}_{Q-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \frac{1}{1-\delta_Q} \begin{bmatrix} \mathbf{y}_Q \mathbf{y}_Q^T & -\mathbf{y}_Q \\ -\mathbf{y}_Q^H & 1 \end{bmatrix}$
- 10: **end for**
- 11: $\tilde{\Psi}^\dagger = \frac{1}{2M} \mathbf{W}_{Q-1}^{-1} \tilde{\Psi}^T$

Pseudoinverse

Iterative calculation of the pseudoinverse (ICP)

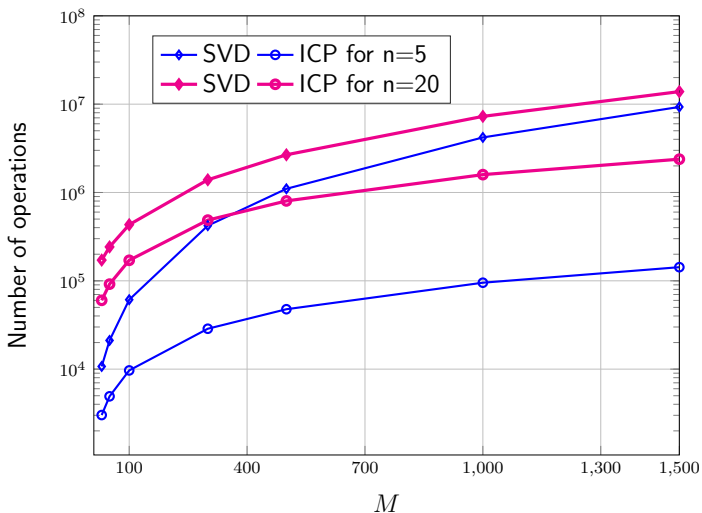


Figure: Operations vs. the number of measurements M

Simulations Results

- The performance is evaluated in terms of the mean square error (MSE)

$$\text{MSE} = \frac{1}{Rn} \sum_{r=1}^R \sum_{i=1}^n |\hat{f}_{i,r} - f_i|^2 \quad (6)$$

between the correct frequencies f_i , $i = 1, 2, \dots, n$ and their estimates $\hat{f}_{i,r}$ in $R = 200$ runs.

- The measurement noise samples are drawn from an i.i.d complex Gaussian random process with zero mean and variance σ^2 .
- All the amplitudes of the complex sinusoids were considered equal to one.
- The number of steps of the VP algorithm was set to 20 for all the test cases.
- We chose the DTFT estimate of the frequencies as initial points of the VP optimization and we assumed that the number of frequencies is known.

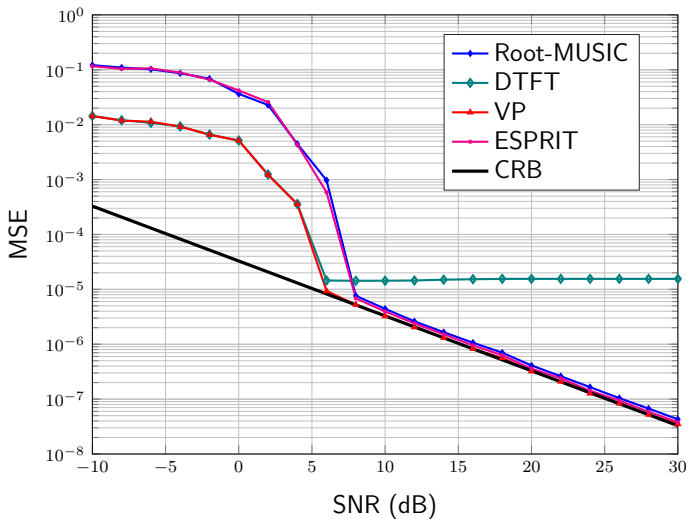


Figure: MSE vs SNR for a scenario with $M = 30$ for five frequencies.

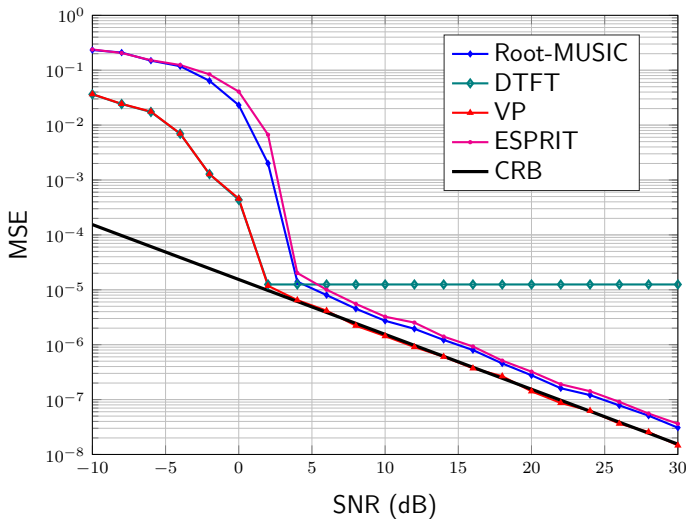


Figure: MSE vs SNR for a scenario with $M = 30$ for two closely spaced frequencies $f_1 = 0.405$, $f_2 = 0.45$.

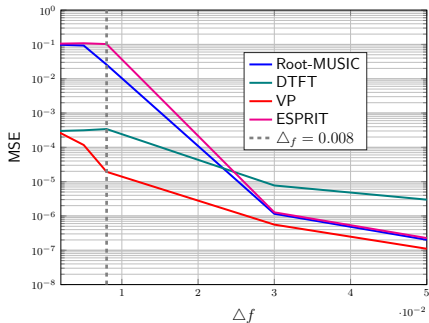
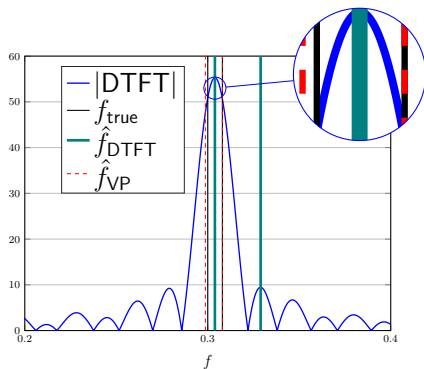
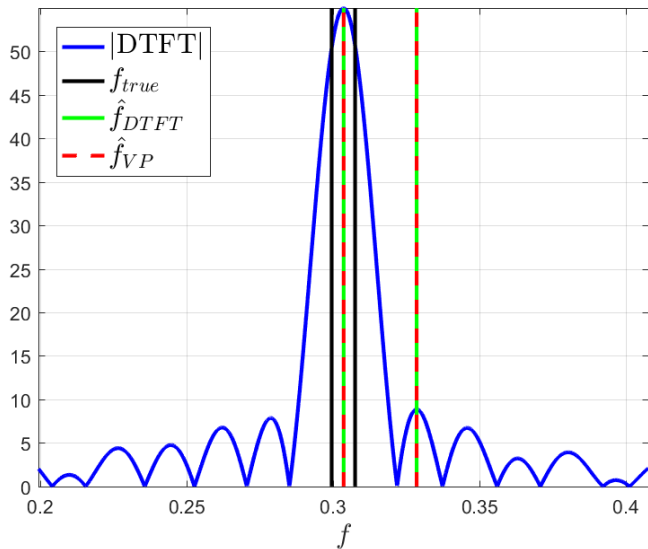
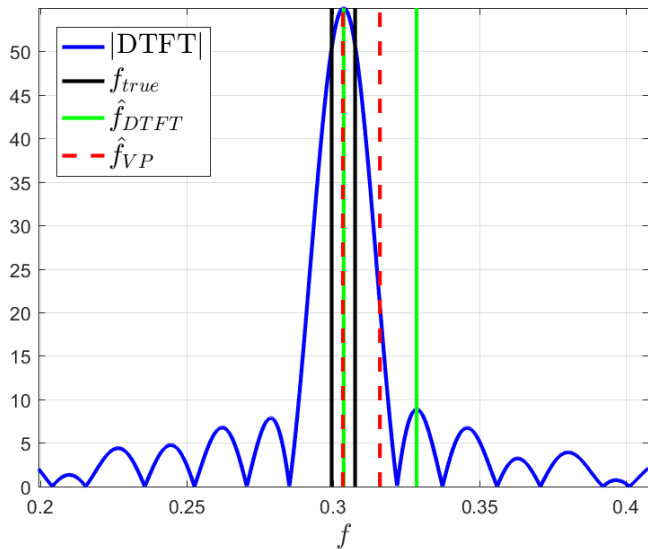
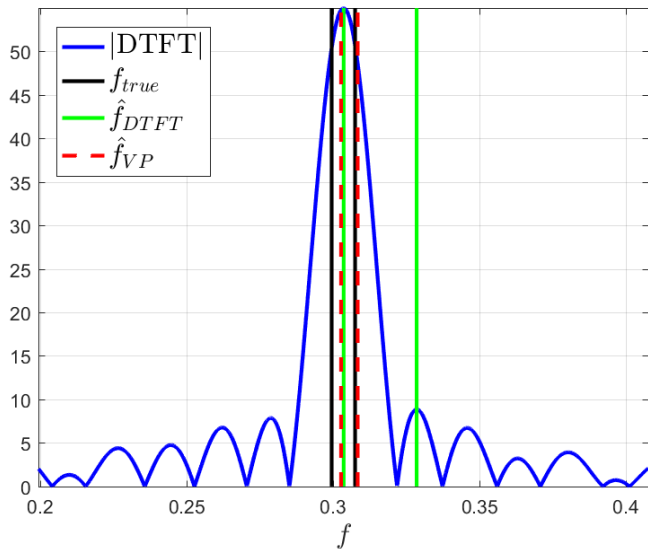
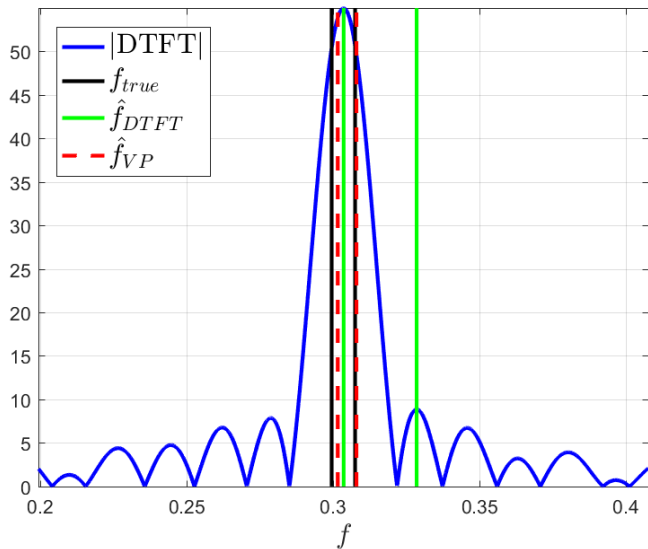
(a) MSE vs. Δf (b) DTFT spectrum for $\Delta f = 0.008$

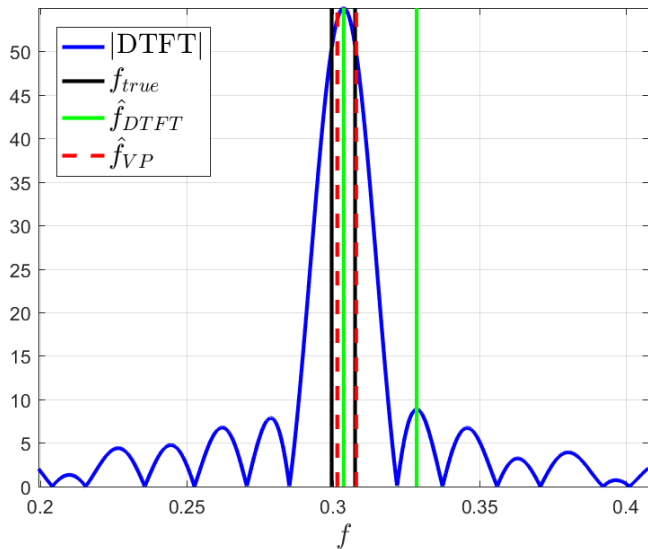
Figure: Analysis of the impact of Δf in the performance, for a scenario $M = 30$, $\text{SNR} = 20\text{dB}$, and two frequencies chosen such that the distance between them is equal to Δf .











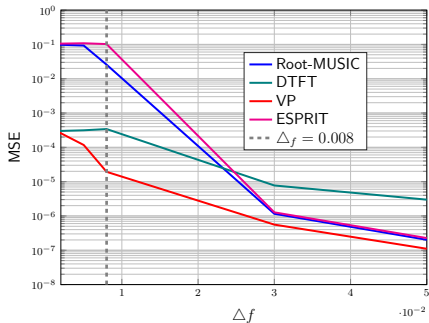
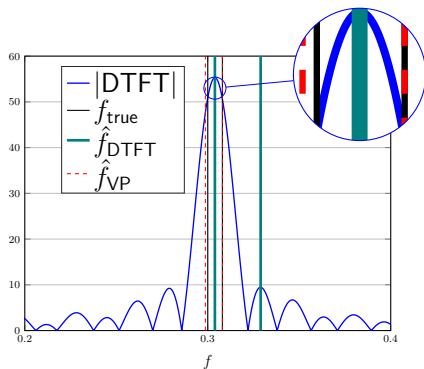
(a) MSE vs. Δf (b) DTFT spectrum for $\Delta f = 0.008$

Figure: Analysis of the impact of Δf in the performance, for a scenario $M = 30$, $\text{SNR} = 20\text{dB}$, and two frequencies chosen such that the distance between them is equal to Δf .

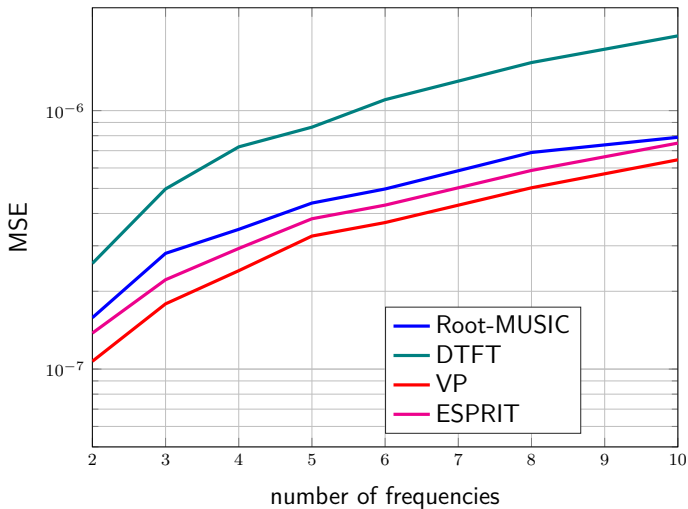


Figure: MSE vs. the number of frequencies for a scenario with $M = 45$ and $\text{SNR}=15\text{dB}$.

Theorem (Jupp, "Lethargy Theorem")

Across the main-faces $s_n^{(p)}$ ($p = 2, \dots, n$) of \bar{s}_n ,

$$\mathbf{n}_p^T \nabla \tilde{F}(\mathbf{f}) = 0,$$

where \mathbf{n}_p is the unit outward normal to $s_n^{(p)}$.

- $s_n = \{\mathbf{f} \in \mathbb{R}^n : 0 < f_1 < f_2 < \dots < f_n < 0.5\}$ represents the parameter space.
- $\bar{s}_n = \{\mathbf{f} \in \mathbb{R}^n : 0 \leq f_1 \leq f_2 \leq \dots \leq f_n \leq 0.5\}$ is the closure of s_n including multiple confluent frequencies

Theorem (Jupp, "Lethargy Theorem")

Across the main-faces $s_n^{(p)}$ ($p = 2, \dots, n$) of \bar{s}_n ,

$$\mathbf{n}_p^T \nabla \tilde{F}(\mathbf{f}) = 0,$$

where \mathbf{n}_p is the unit outward normal to $s_n^{(p)}$.

The lethargy theorem has the following consequences:

- the reduced and so the full functional has many stationary points on the main-faces of \bar{s}_n .
- $\tilde{F}(\mathbf{f})$ and $F(\mathbf{c}, \mathbf{f})$ are non-convex for any set of data, and any choice of smooth convex norm.
- numerical optimizers have poor convergence near the boundary of s_n .

D. L. B. Jupp, "The Lethargy Theorem – A Property of Approximation by γ -Polynomials," *Journal of Approximation Theory*, vol. 14, pp. 204–217, 1975.

D. L. B. Jupp, "Approximation to data by splines with free knots," *SIAM Journal on Numerical Analysis*, vol. 15, no. 2, pp. 328–343, 1978.

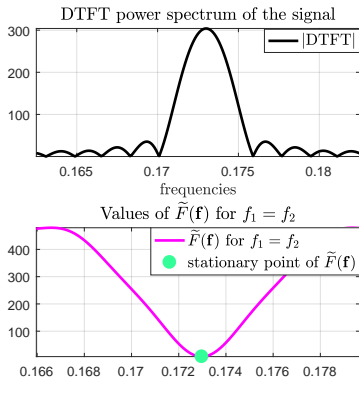
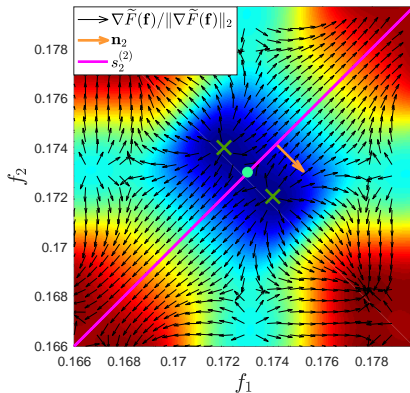
(a) Extreme points on $s_2^{(2)}$.(b) Graph of $\tilde{F}(\mathbf{f})$ for $\mathbf{f} \in \mathbb{R}^2$.

Figure: Values of $\tilde{F}(\mathbf{f})$, where the signal contains two normalized frequency components $f_1 = 0.171$ and $f_2 = 0.174$.

Conclusions

- We have formulated the frequency estimation problem as an SNLLS problem.
- We quantified the difficulty of the corresponding optimization problem by applying a lethargy-type theorem.
- Based on this representation, we have proposed a VP optimization for finding the frequency parameters.
- An efficient way of calculating the exact gradient of the VP functional has been presented, with a lower computational cost than existing techniques.
- Simulations have shown that the proposed estimator outperforms previously reported techniques in scenarios with closely spaced frequencies and achieves more accurate results in terms of the MSE.
- The interactive version of the code is available at <https://codeocean.com/capsule/5263510/tree/v1>

Thanks!

yuneisy.garcia_guzman@jku.at