Generalized autocorrelation analysis for multi-target detection

Ye'Ela Shalit[1], Ran Weber[1], Asaf Abas[2], **Shay Kreymer**[1], and Tamir Bendory[1]

[1]School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

[2]Department of Applied Mathematics, Tel Aviv University, Tel Aviv, Israel

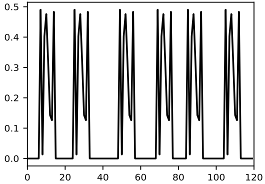
ICASSP 2022, May 25, 2022

Multi-target detection (MTD)

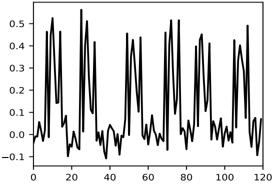
The problem of estimating a target signal $x \in \mathbb{R}^L$ from a noisy measurement $y \in \mathbb{R}^N$ that contains multiple copies of the signal, each randomly translated:

$$y[\ell] = \sum_{i=1}^{p} x[\ell - \ell_i] + \varepsilon[\ell], \tag{1}$$

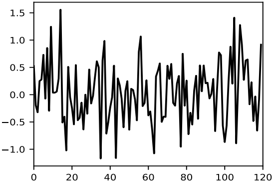
where $\{\ell_i\}_{i=1}^{p} \in \{L+1, \ldots, N-L\}$ and $\varepsilon[\ell] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.



(a) No noise.



(b) SNR = 50.



(c) SNR = 0.1.

Cryo-electron microscopy (cryo-EM)

Single-particle cryo-electron microscopy (cryo-EM) is an emerging technology for macromolecular structure determination.

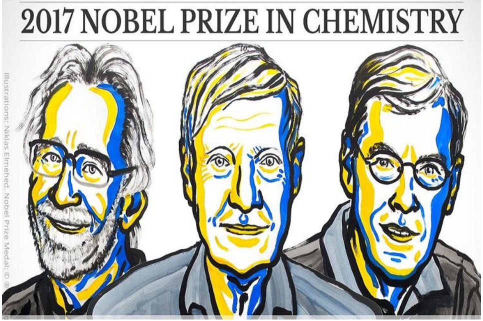


Figure: The Nobel Prize in Chemistry 2017 was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution".

Cryo-EM

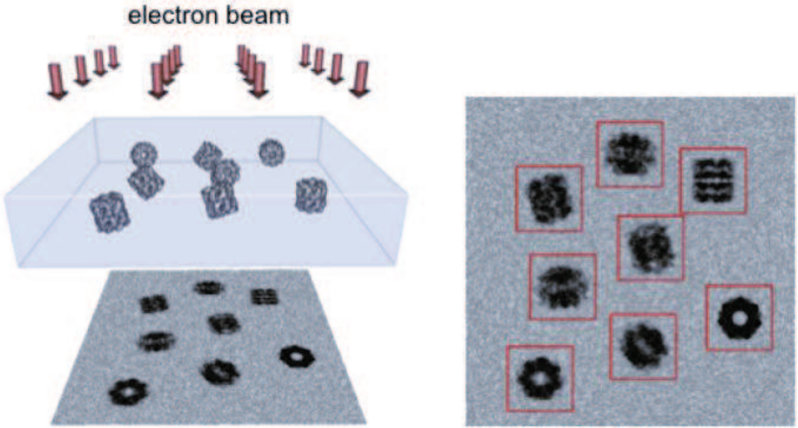


Figure: The cryo-EM process and particle picking.

Autocorrelation analysis

The idea: **finding a signal that best explains the empirical autocorrelations of the measurement**:

$$A_y^1 := \frac{1}{N} \sum_{i \in \mathbb{Z}} y[i], \tag{2}$$

$$A_y^2[\ell_1] := \frac{1}{N} \sum_{i \in \mathbb{Z}} y[i] y[i+\ell_1], \tag{3}$$

$$A_y^3[\ell_1, \ell_2] := \frac{1}{N} \sum_{i \in \mathbb{Z}} y[i] y[i+\ell_1] y[i+\ell_2]. \tag{4}$$

Autocorrelation analysis

For any fixed level of noise σ^2, density γ and signal length L, in the limit $N \to \infty$, we have that:

$$A_y^1 \overset{\text{a.s.}}{=} \gamma A_x^1, \tag{5}$$

$$A_y^2[\ell_1] \overset{\text{a.s.}}{=} \gamma A_x^2[\ell_1] + \sigma^2 \delta[\ell_1], \tag{6}$$

$$A_y^3[\ell_1, \ell_2] \overset{\text{a.s.}}{=} \gamma A_x^3[\ell_1, \ell_2] + \gamma A_x^1 \sigma^2 (\delta[\ell_1] + \delta[\ell_2] + \delta[\ell_1 - \ell_2]), \tag{7}$$

for $\ell_1, \ell_2 \in \mathcal{L}$, where δ is the Kronecker delta function.
Here, γ is the density of the target signals in the measurement and is defined by

$$\gamma = p \frac{L}{N}. \tag{8}$$

The autocorrelations of x and y do not directly depend on the location of individual signal occurrences in the measurement, but only through the density parameter γ.

Autocorrelation analysis

In the standard approach, we find the signal that best matches the observable autocorrelations by minimizing a LS objective:

$$\underset{x\in\mathbb{R}^L,\gamma>0}{\arg\min}\,(A_y^1-\gamma A_x^1)^2+w_2\sum_{\ell_1=0}^{L-1}\|A_y^2[\ell_1]-\gamma A_x^2[\ell_1]-\sigma^2\delta[\ell_1]\|_2^2$$
$$+w_3\sum_{\ell_1=0}^{L-1}\sum_{\ell_2=0}^{L-1}\|A_y^3[\ell_1,\ell_2]-\gamma A_x^3[\ell_1,\ell_2]-\gamma A_x^1\sigma^2(\delta[\ell_1]+\delta[\ell_2]+\delta[\ell_1-\ell_2])\|_2^2, \tag{9}$$

where the weights w_2 and w_3 were chosen such that each term is equally weighted.

Generalized autocorrelation analysis

Generalized autocorrelation analysis consists of replacing the LS objective function by weighted LS with optimal weights.

A specific choice of weights guarantees favorable asymptotic statistical properties, such as minimal asymptotic variance of the estimation error [Hansen 1982].

The **moment function** $f(\theta, y)$ for $\theta \in \Theta$ is chosen such that its expectation is zero only at a single point $\theta = \theta_0$, where θ_0 is the ground truth parameter (in our case, the target signal x and the density parameter γ). Namely,

$$\mathbb{E}\left[f(\theta, y)\right] = 0 \quad \text{if and only if} \quad \theta = \theta_0. \tag{10}$$

Generalized autocorrelation analysis

For MTD, we define the i-th observation from the measurement y

$$y_i := [y[i], \dots, y[i+L-1]] \in \mathbb{R}^L, \tag{11}$$

and the moment function $f(\theta, y_i)$ simply as the discrepancy between the autocorrelations of y_i and the population autocorrelations.

The estimated sample moment function is the average of f over N observations:

$$g_N(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} f(\theta, y_i). \tag{12}$$

The generalized autocorrelation estimator is defined as the minimizer of the weighted LS objective

$$\hat{\theta}_N = \arg \min_{\theta \in \Theta} \; g_N(\theta)^T W_N g_N(\theta). \tag{13}$$

Generalized autocorrelation analysis

Let

$$S := \lim_{N \to \infty} \mathrm{Cov}\left[\sqrt{N} g_N(\theta_0)\right], \tag{14}$$

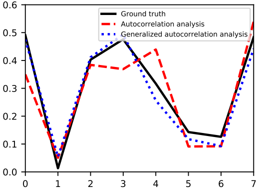
be the covariance matrix of the estimated sample moment function (12) at the ground truth θ_0.

It can be shown that the optimal choice of a weighting matrix is given by

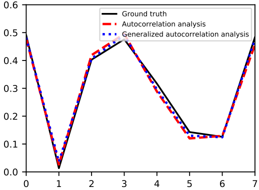
$$W = S^{-1}. \tag{15}$$

Numerical experiments[1]

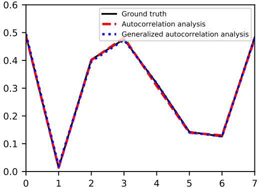
Recovery from a noisy measurement



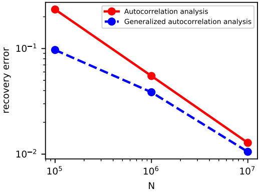
(a) $N = 10^5$



(b) $N = 10^6$



(c) $N = 10^7$



(d) Recovery error

[1] The code to reproduce all experiments is publicly available at https://github.com/krshay/MTD-GMM.

Numerical experiments

Recovery error as a function of the measurement length

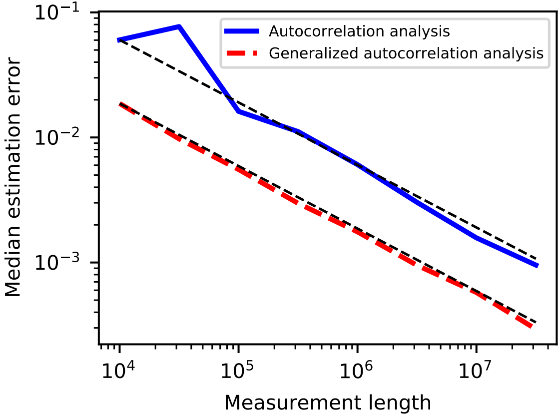


Figure: The median estimation error as a function of the measurement length N with $\mathrm{SNR} = 50$, by: (a) autocorrelation analysis; (b) generalized autocorrelation analysis. The black dashed lines illustrates a slope of $-1/2$, as predicted by the law of large numbers.

Numerical experiments

Recovery error as a function of the SNR

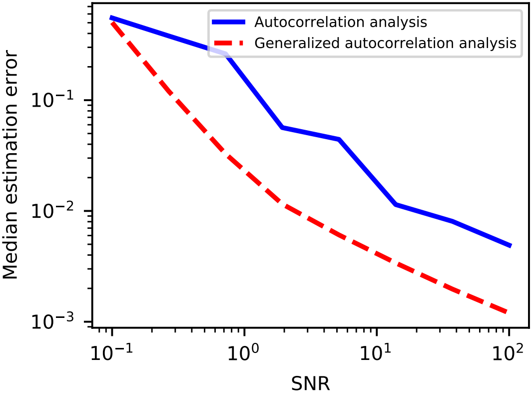


Figure: The median estimation error as a function of SNR, for measurements with length $N = 10^6$, by: (a) autocorrelation analysis; (b) generalized autocorrelation analysis estimator. Evidently, the generalized autocorrelation analysis estimator outperforms classical autocorrelation analysis for all SNR levels.

Conclusion

- The main contribution of this study is incorporating the generalized method of moments into the computational framework for the MTD problem.
- We demonstrate a successful signal reconstruction directly from noisy measurements.
- The generalized autocorrelation analysis framework outperforms autocorrelation analysis in a wide range of parameters.
- A step towards efficiently estimating a molecular structure directly from a noisy cryo-EM micrograph.

Thanks for your attention!