Structural Optimization of Multi-chromophore Fluorescence Detection Experiments.

Alexei Goun
Department of Chemistry, Princeton University,
Frick Laboratory Bldg, Washington Rd, Princeton NJ, 08540
agoun@princeton.edu

Abstract: Fluorescent protein chromophores are the workhorse of molecular biology. Multi-chromophore mixture composition detection is a challenging problem in fluorescent spectroscopy and microscopy. We developed a predictive model of the performance based on Cramer-Rao uncertainty bound and its representation through Cauchy-Binet equation. Optimal design of detection experiment is presented.

1. Introduction.

Introduction of fluorescent proteins revolutionized the field of molecular biology by allowing the observation of intracellular processes at a single protein level [1]. At the same time detection of such a small signal is a challenging problem of statistical inference.

In past few years the range of single molecule techniques expanded greatly, techniques such as hyperspectral imaging [2], lifetime imaging [3], selective excitation with tunable lasers [4], and other approaches are commonly used. Most of the single of few molecule measurements are unique, require considerable investment of time and resources. Moreover they can be accomplished with a wide range of techniques or combination of techniques each providing its own performance. From experimental point of view it would be highly desirable to estimate the performance of the system before the experimental design is implemented.

The approach, based on the Cramer-Rao [5, 6] bound offers a very flexible framework where different experimental designs can be compared and optimized.

2. Cramer-Rao Bound, Fisher Information, and Cauchy-Binet Representation.

Let's consider a detection system with *k* detection channels, which is utilized to detect the composition of the mixture of *p* fluorescent proteins. Experimental observations are described by the probability density function of photocounts.

$$P = P(n_1, n_2, \dots, n_k; m_1, m_2, \dots, m_p) \quad (1)$$

The average number of photocounts is in linear dependence to the concentration of fluorescent proteins $n_i = \sum_{j=1}^{p} \alpha_{ij} m_j + n_i^N$; α_{ij} is the number of photons in the detection channel *i*, produced by the fluorescent protein *j*, n_i^N is the total number of the noise photocounts in the channel *i*.

The Cramer-Rao inequality provides the following bound on the variance of the estimator of fluorescent protein concentrations *m*: $var(\hat{m}) \geq \frac{1}{I(\hat{m})}$, where $I(m)$ is the Fisher information matrix with elements given by

$$I_{ij} = -E\left[\frac{\partial^2 log\left(P(n_1,n_2,\dots,n_k;m_1,m_2,\dots,m_p)\right)}{\partial m_i \partial m_j}\right]. \quad (2)$$

For the Poissonian distribution of photocounts in detection channels, the Fisher information matrix can be represented in the following form $I_{ij} = A^T A$, where a rectangular matrix *A* describes available detection channels.

$$A = \begin{pmatrix} \frac{\alpha_{11}}{\sqrt{n_1}} & \frac{\alpha_{12}}{\sqrt{n_1}} & \cdots & \frac{\alpha_{1m}}{\sqrt{n_1}} \\ \frac{\alpha_{21}}{\sqrt{n_2}} & \frac{\alpha_{22}}{\sqrt{n_2}} & \cdots & \frac{\alpha_{2m}}{\sqrt{n_2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\alpha_{k1}}{\sqrt{n_k}} & \frac{\alpha_{k2}}{\sqrt{n_k}} & \cdots & \frac{\alpha_{km}}{\sqrt{n_k}} \end{pmatrix} \quad (3)$$

The inverse of Fisher information matrix determinant $|I(m)|$ determines the averaged uncertainty of the composition measurement. Consequently, in order to design the optimal detection experiment, one needs to increase the value of the determinant.

Cauchy-Binet equation provides highly illuminating and convenient representation of the determinant of the matrix $I(m)$. This equation states that the one needs to form all possible square submatrices out of the rectangular matrix *A* and sum all their squared determinants.

$$|I| = \sum_{S\in\binom{[p]}{k}} \left|A_{[p],k}\right|^2 \qquad (4)$$

3. Computational Example.

It is easy to see that each square submatrix represents a possible experimental configuration, sufficient to determine the composition of fluorescent protein mixture. Consequently in order to find the optimal experimental configuration one needs to find a square submatrix of *A* with largest value of the determinant. This approach is illustrated on the Figures 1 and 2.

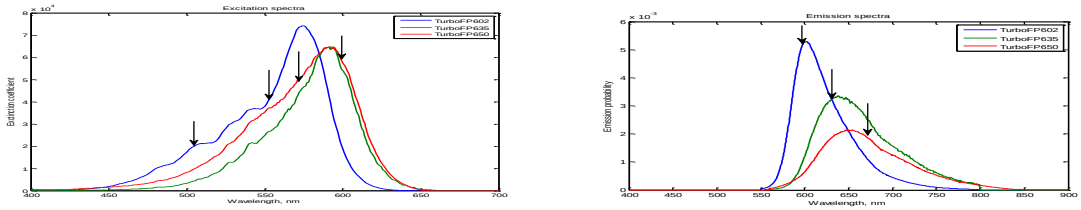

Figure 1. a. Absorption spectra of fluorescent proteins, arrows point to chosen excitation wavelengths. b. Emission spectra of proteins, arrows point to centers of bandpass filters.

Let us consider the case of composition detection of 3 fluorescent proteins: FP602, FP635, and FP650. Suppose the excitation is carried out at 4 different wavelengths: 500, 560, 574, and 600nm. The detection is carried out through 3 different 20nm badndpass filters centered at 590, 620, and 670nm. The combination of the single excitation wavelength with a single bandpass filter provides 12 possible detection channel. The combination of 3 detection channels is sufficient for the determination of the composition of the mixture. There are 351possible experimental configurations.

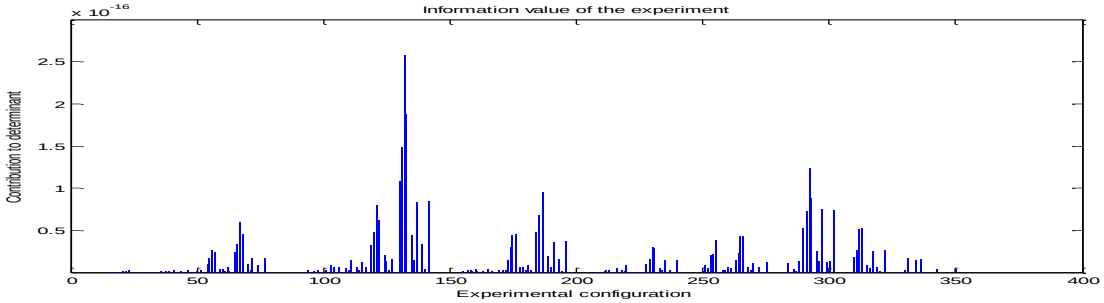

Figure 3. Information content of possible experimental configurations.

The search over the space of experimental configurations provides the following experimental assignment of excitation/detection channels 574nm/569nm, 600nm/620nm, 560nm/670nm.

4. References.

[1] Nathan C Shaner, Paul A Steinbach, Roger Y Tsien, "A guide to choosing fluorescent proteins," Nature Methods **2**, 905 - 909 (2005).

[2] Vincent Studer, et al, "Compressive fluorescence microscopy for biological and hyperspectral imaging," *PNAS* **109**, E1679-E1687 (2012).

[3] Philip Tinnerfeld, et al, "Confocal Fluorescence Lifetime Imaging Microscopy (FLIM) at the Single Molecule Level," Single Molecules **1**, 215-223 (2000).

[4] George H. Patterson, Jennifer Lippincott-Schwartz, "A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells," Science **297**, 1873-1877 (2002).

[5] Rao, Calyampudi Radakrishna "Information and the accuracy attainable in the estimation of statistical parameters". Bulletin of the Calcutta Mathematical Society. 37: 81–89 (1945).

[6] Harald Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: (Princeton Univ. Press. 1946).