Initial Guesses Generation for Fluorescence Intensity Distribution Analysis

Victor V. Skakun[1,3], Eugene G. Novikov[2],. Vladimir V. Apanasovich[3], Hans Tanke[4], Andre Deelder[1], Oleg A. Mayboroda[1]
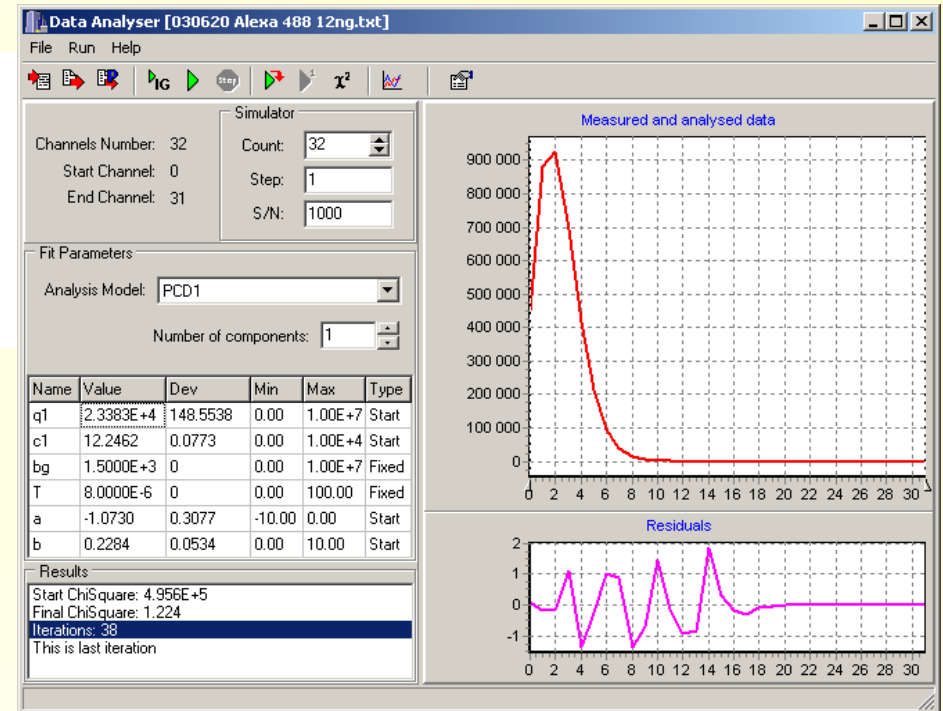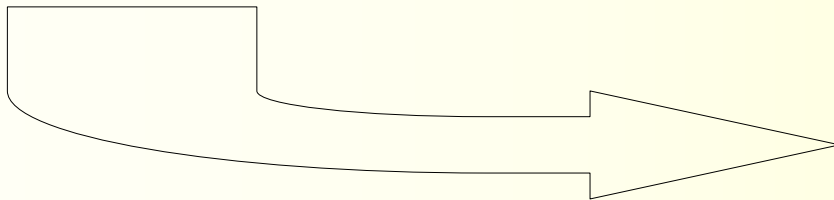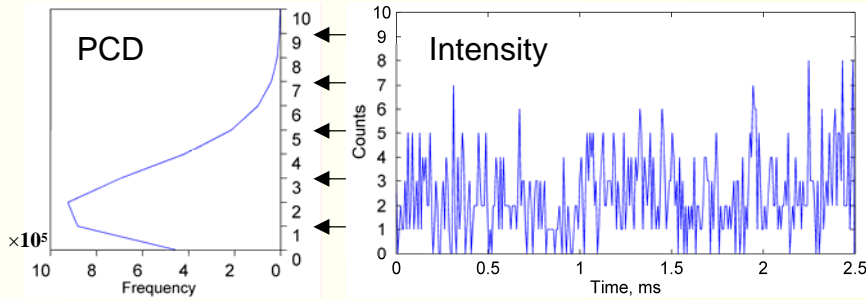
[1] Department of Parasitology, Leiden University Medical Centre, Leiden, The Netherlands

[2] Service Bioinformatique, Institut CURIE, Paris, France

[3] Department of Systems Analysis, Belarusian State University, Minsk, Belarus

[4] Department of Molecular Cell Biology, Leiden University Medical Centre, Leiden, The Netherlands

Why IG?

IG generation

Allows to:

➢ increase efficiency and correctness of the fit

➢ avoid trapping into local minima

➢ increase robustness of the analysis

➢ decrease number of iterations

➢ reduce user participation and make the whole procedure more standardized

Demands:

➢ developing of straightforward, noniterative methods

Method of moments

$$M_k(\eta_1, \eta_2, \ldots, \eta_m) = \tilde{M}_k, \quad k = 1, 2, \ldots, m, \quad (1)$$

where $\eta_1, \eta_2, \ldots, \eta_m$ is a set of unknown parameters

$$\tilde{M}_k = <n^k> = \sum_{n=1}^{N-1} n^k P^*(n) \quad (2)$$

$P^*(n)$ is a probability to get n photons within a counting time interval T

in application to factorial cumulants

$$K_k(\eta_1, \eta_2, \ldots, \eta_m) = \tilde{K}_k \quad (3)$$

$$K_k = \left. \frac{d^k \ln G(\xi)}{d\xi^k} \right|_{\xi=1} \quad (4) \qquad G(\xi) = \sum_{n=0}^{\infty} \xi^n P(n) \quad (5)$$

$$\tilde{F}_k = <n(n-1)\ldots(n-k+1)> = \sum_{n=k}^{N-1} n(n-1)\ldots(n-k+1) P^*(n) \quad (6)$$

$$\tilde{K}_k = \tilde{F}_k - \sum_{i=1}^{k-1} \binom{k-1}{i} \tilde{K}_{k-i} \tilde{F}_i, \quad \text{where} \quad \binom{k-1}{i} = \frac{(k-1)!}{i!(k-i-1)!} \quad (7)$$

Historical background

Moment analysis of fluorescence fluctuations

$$\begin{cases} <\Phi> = \chi_1 \sum_i q_i c_i \\ <\Delta\Phi^2> = \chi_2 \sum_i q_i^2 c_i \\ <\Delta\Phi^3> = \chi_3 \sum_i q_i^3 c_i \\ <\Delta\Phi^4> - 3<\Delta\Phi^2>^2 = \chi_4 \sum_i q_i^4 c_i \end{cases} \quad (8)$$

here Φ is the fluorescence intensity, c_i is the number of molecules per observation volume, q_i is the specific brightness expressed in cpm, $B(r)$ is a spatial brightness function, i is the number of molecular species
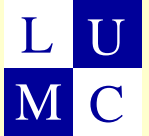
$$\chi_k = \int B^k(r)\,dr \quad (9)$$

Qian, H., and E.L. Elson. Biophys. J. 57, 1990
Qian, H., and E.L. Elson. PNAS 87, 1990

Fluorescence Cumulant Analysis

$$K_k = \chi_k \sum_i c_i q_i^k \quad (10)$$

Muller J.D. Biophis. J. 86, 2004

FIDA

$$G(\xi) = \sum_{n=0}^{\infty} \xi^n P(n) \quad (11) \qquad P(n) \text{ is photon counting distribution (PCD)}$$

$$G(\xi) = \exp\left((\xi-1)\lambda T + \sum_j c_j \int_V \left\{ \exp\left[(\xi-1)q_j TB(r) \right] - 1 \right\} dV \right) \quad (12)$$

here c_j is the mean number of molecules per observation volume, q_j is the specific brightness expressed in cpmt, V is the observation volume, T is the counting time interval, $B(r)$ is brightness profile function which is the product excitation intensity and detection efficiency, j is the number of molecular species and λ is the mean background count rate of detector.

Evotec Biosystems AG. 1998. Int. Patent WO 98/16814.

Kask at al. PNAS 96, 1999.

$$\frac{dV}{dx} = A_0 \ (x + ax^2 + bx^3), \quad x = \ln\left[B_0 \Big/ B(r) \right], \quad B(r) = B_0 \, e^{-x} \quad (13)$$

where a, b are instrumental parameters
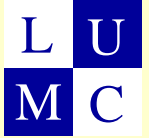and A_0, B_0 can be calculated from system of normalization equations:

$$\chi_1 = \int_V B(r) dV = 1,$$

$$\quad (14) \qquad \text{Finally} \quad P(n) = FFT^{-1}(G(e^{i\varphi}))$$

$$\chi_2 = \int_V B^2(r) dV = 1.$$

Evotec Biosystems AG. 1998. Int. Patent WO 98/16814.

Palo at al. Boiphis. J. 79, 2000

General system of equations for IG generation

$$K_k = \left. \frac{d^k \ln G(\xi)}{d\xi^k} \right|_{\xi=1}$$

$$K_1 = (\lambda + \sum_j c_j q_j)T$$

$$K_2 = \sum_j c_j q_j^2 T^2 \qquad\qquad (15)$$

$$K_k = \chi_k \sum_j c_j q_j^k T^k, \quad k = 3, 4, \ldots,$$

$$\chi_k = \int_0^\infty (B_0 e^{-x})^k A_0 (x + ax^2 + bx^3)\,dx \qquad (16)$$

$$B_0 = \frac{8(2a + 6b + 1)}{2a + 3b + 2}, \quad A_0 = \frac{2a + 3b + 2}{8(2a + 6b + 1)^2} \qquad (17)$$

IG for one component model

Basic system of equations:

estimated parameters are c, q, λ, a, b

$$\begin{cases} K_1 = (\lambda + cq)T \\ K_2 = cq^2 T^2 \\ K_3 = cq^3 T^3 \dfrac{64(2a + 6b + 1)(2a + 2b + 3)}{27(2a + 3b + 2)^2} \\ K_4 = cq^4 T^4 \dfrac{4(2a + 6b + 1)^2(4a + 3b + 8)}{(2a + 3b + 2)^3} \\ K_5 = cq^5 T^5 \dfrac{4096(2a + 6b + 1)^3(10a + 6b + 25)}{625(2a + 3b + 2)^4} \end{cases} \quad (18)$$

Solution:

$$\begin{cases} \dfrac{(10a + 6b + 25)(2a + 2b + 3)}{(4a + 3b + 8)^2} = \dfrac{16875\, K_5 K_3}{16384\, K_4^2} \\ \dfrac{729(2a + 3b + 2)(4a + 3b + 8)}{(2a + 2b + 3)^2} = \dfrac{1024\, K_4 K_2}{729\, K_3^2} \end{cases} \quad (19)$$

$$\lambda = K_1 - \frac{64\, K_1^2}{27\, K_2} \frac{(2a + 2b + 3)(2a + 6b + 1)}{(2a + 3b + 2)^2} \quad (20)$$

$$q = \frac{K_2}{(K_1 - \lambda T)T}, \quad c = \frac{(K_1 - \lambda T)^2}{K_2} \quad (21)$$

Simplifications: background (λ) is known

estimated parameters are c, q, a, b

$$\begin{cases} \chi_3 = \dfrac{(K_1 - \lambda T)K_3}{K_2^2} = \dfrac{64(2a + 6b + 1)(2a + 2b + 3)}{27(2a + 3b + 2)^2} \\ \chi_4 = \dfrac{(K_1 - \lambda T)^2 K_4}{K_2^3} = \dfrac{4(2a + 6b + 1)^2(4a + 3b + 8)}{(2a + 3b + 2)^3} \end{cases} \quad (22)$$

$$q = \frac{K_2}{(K_1 - \lambda T)T}, \quad c = \frac{(K_1 - \lambda T)^2}{K_2} \quad (21)$$

In general $\quad \chi_k = \dfrac{(K_1 - \lambda T)^{k-2} K_k}{K_2^{k-1}} \quad (23)$

Solution of system 19

Shape of the fourth order polynomial with respect to parameter *a*

$$H_1 a^4 + H_2 a^3 + H_3 a^2 + H_4 a + H_5 = 0$$

$$\alpha^2 \beta^2 - 18\alpha\beta + 27\alpha + 16\beta - 27 > 0 \quad - \quad \text{red line}$$

$$\alpha^2 \beta^2 - 18\alpha\beta + 27\alpha + 16\beta - 27 \leq 0 \quad - \quad \text{blue line}$$

$$\alpha = \frac{16875 \, K_5 K_3}{16384 \, K_4^2} \qquad \beta = \frac{1024 \, K_4 K_2}{729 \, K_3^2}$$

Root selection:

➢ setting admissible ranges
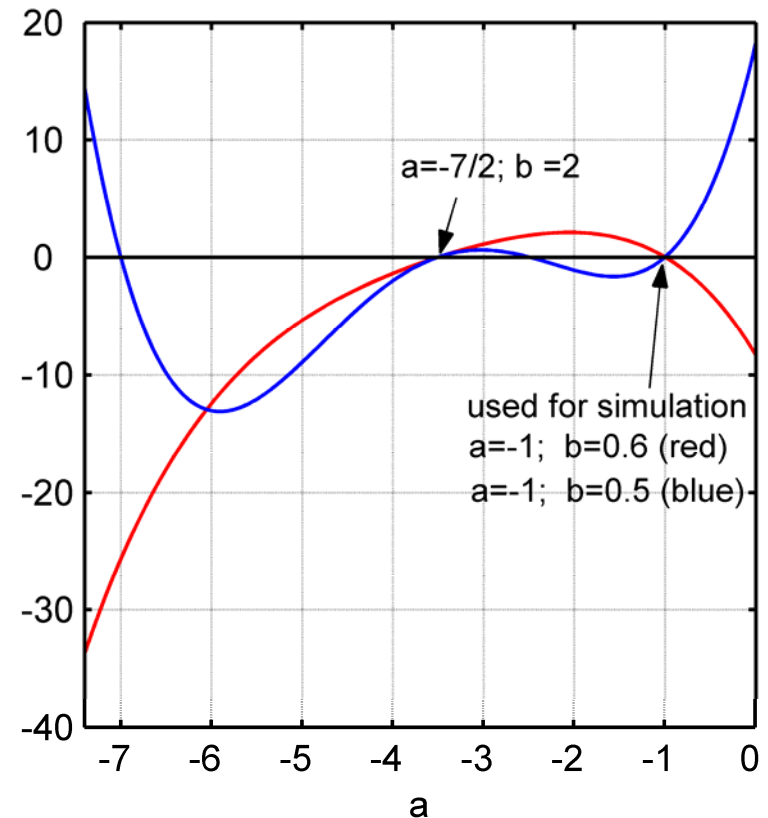➢ minimization of χ^2 criterion

Fig. 1. Roots of polynomial

IG for two component model

Initial system of equations

estimated parameters are c_1, c_2, q_1, q_2, λ, a, b

$$K_1 = (\lambda + c_1 q_1 + c_2 q_2)T$$

$$K_2 = (c_1 q_1^2 + c_2 q_2^2)T^2$$

$$K_3 = (c_1 q_1^3 + c_2 q_2^3)T^3 \frac{64(2a + 2b + 3)(2a + 6b + 1)}{27(2a + 3b + 2)^2}$$

$$K_4 = (c_1 q_1^4 + c_2 q_2^4)T^4 \frac{4(4a + 3b + 8)(2a + 6b + 1)^2}{(2a + 3b + 2)^3} \quad (24)$$

$$K_5 = (c_1 q_1^5 + c_2 q_2^5)T^5 \frac{4096(10a + 6b + 25)(2a + 6b + 1)^3}{625(2a + 3b + 2)^4}$$

$$K_6 = (c_1 q_1^6 + c_2 q_2^6)T^6 \frac{4096(2a + b + 6)(2a + 6b + 1)^4}{27(2a + 3b + 2)^5}$$

$$K_7 = (c_1 q_1^7 + c_2 q_2^7)T^7 \frac{8^6(14a + 6b + 49)(2a + 6b + 1)^5}{2401(2a + 3b + 2)^6}$$

Simplifications:

1. background (λ) and instrumental parameters a and b are known

$$K_1 - \lambda T = (c_1 q_1 + c_2 q_2)T$$

$$K_2 = (c_1 q_1^2 + c_2 q_2^2)T^2$$

$$K_3 / \chi_3 = (c_1 q_1^3 + c_2 q_2^3)T^3 \quad (25)$$

$$K_4 / \chi_4 = (c_1 q_1^4 + c_2 q_2^4)T^4,$$

$$\chi_3 = \frac{64(2a + 6b + 1)(2a + 2b + 3)}{27(2a + 3b + 2)^2}$$

$$\chi_4 = \frac{4(2a + 6b + 1)^2(4a + 3b + 8)}{(2a + 3b + 2)^3} \quad (26)$$

2. background (λ) is known

A number of predefined parameters a and b used for solution of the sysmem 25. As result a number of sets of parameters is generated and the set resulting in lowest χ^2 criterion is accepted.

Testing of IG for one component model on simulated data
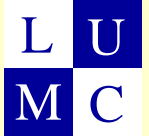
1. IG for all parameters *c*, *q*, λ, *a*, *b*

Table 1. IG for one component model on noisy PCD at different S/N. IG were rejected when either λ was negative or *a* and *b* exceed bounds (-2, 0); (0, 2) respectively. $T = 5\times10^{-5}$. 50 simulations in each series.

Parameter	Used for simulation	Recovered		
		$S/N_i = 7000$	$S/N_i = 3000$	$S/N_i = 1000$
c	5	4.994±0.126	4.882±0.205	4.327±0.542
q	20000	20016±252	20253±436	21633±1462
λ	2000	2064±1260	3208±2086	9170±5948
a	-1	-0.999±0.019	-0.980±0.034	-0.853±0.152
b	0.5	0.500±0.002	0.499±0.003	0.495±0.009

$$S/N = \sqrt{mp_{max}/(1-p_{max})}$$

$$S/N_{initial} = \sqrt{mp_{max}} =$$

$$= \sqrt{\text{Value at Maximum}}$$

here $p_{max} = \max_n(P(n))$,

m is total number of photons

2. IG for parameters *c*, *q*, *a*, *b* (λ is known)

Table 2. IG calculated for one component model on noisy PCD at different S/N. λ fixed to 2000. $T = 5\times10^{-5}$.

Parameter	Used for simulation	Recovered		
		$S/N_i = 1000$	$S/N_i = 300$	$S/N_i = 50$
c	5	5.000±0.006	4.999±0.025	4.999±0.107
q	20000	20002±24	20003±107	20029±455
a	-1	-1.000±0.004	-1.006±0.015	-1.048±0.106
b	0.5	0.500±0.008	0.514±0.033	0.522±0.183

# Testing of IG for two component model on simulated data

IG for parameters c_1, c_2, q_1, q_2 (λ, a, b are known)

Table 3. IG calculated for two component model on noisy PCD at different S/N. $\lambda = 1000$. $a = -1$; $b = 0.5$; $T = 2 \times 10^{-5}$. 50 simulations in each series.

Parameter	Used for simulation	Recovered	
		$S/N_i = 1000$	$S/N_i = 100$
c_1	10	9.99±0.13	10.11±1.70
q_2	20000	19883±635	18282±5013
c_1	2	2.04±0.23	2.59±1.64
q_2	50000	49847±1404	50931±11676

Testing of IG for one component model on measured data
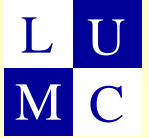
1. IG for all parameters c, q, λ, a, b

Table 4. IG calculated for one component model on measured data (Alexa 488). $T = 8\times10^{-6}$.

Parameters	IG	Fit starting from IG
χ^2	0.781	0.760
c	23.12	25.08±4.07
q	19192	18426±1495
λ	18424	27±37465
a	-4.41	-4.58±0.47
b	3.22	3.47±0.63

Confidential intervals are calculated as Asymptotic Standard Errors.

2. IG for parameters c, q, a, b (λ is estimated from additional measurement)

Table 5. IG calculated for one component model on measured data (Alexa 488). λ fixed to 1500. $T = 8\times10^{-6}$

Parameters	IG	Best fit = fit starting from IG
χ^2	0.668	0.666
c	24.916	24.923±0.189
q	18500	18500±140.5
a	-1.463	-1.472±0.0014
b	0.321	0.324±0.0004

Confidential intervals are calculated as Asymptotic Standard Errors.
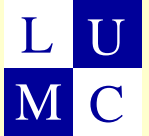
Testing of IG for two component model on measured data

L U
M C

Table 8. IG calculated for two component model on measured data (mixture of IgG labeled with Alexa 488 and pure dye). λ fixed to 1000. $T = 2 \times 10^{-5}$

Parameters	IG	Best fit
x^2	1.26	0.78
c_1	5.477	5.166±0.089
q_1	32099	32768±872
c_2	0.387	0.528±0.052
q_2	90770	78824±2511
λ	1000 (fixed)	1000 (fixed)
a	-0.85	-0.769±0.014
b	0.25	0.296±0.018

Confidential intervals are calculated as Asymptotic Standard Errors.

minima in χ^2 space

Table 6. Parameters used for calculation
(obtained from best fit of Alexa 488). $T = 8 \times 10^{-6}$.

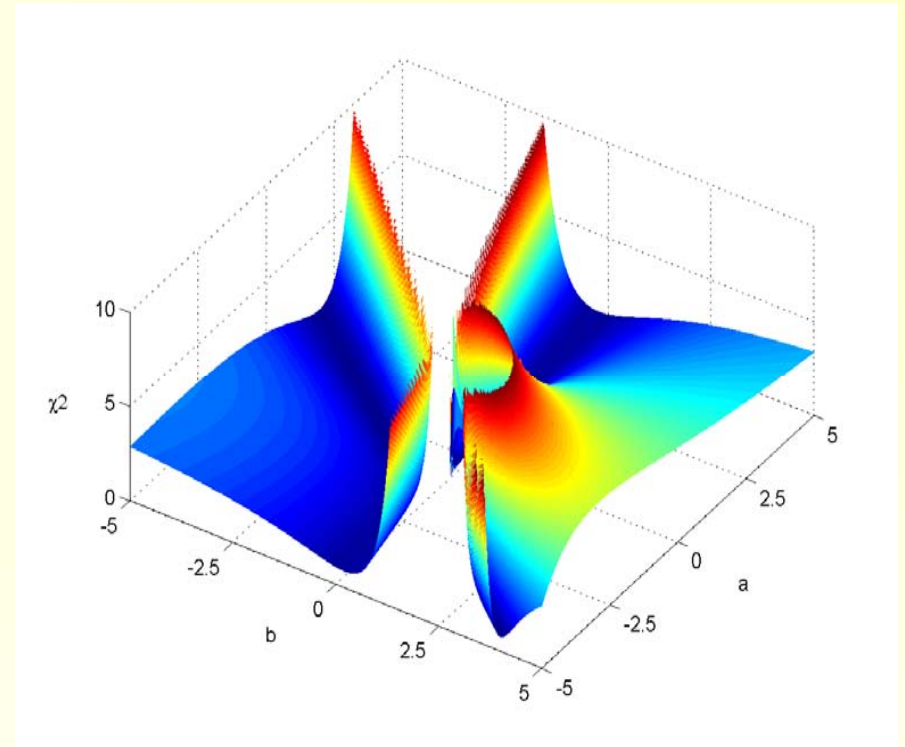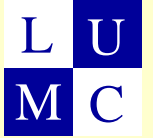| Parameter	Value
c	24.923
q	18500
λ	1500
a	varied
b	varied

Fig. 2. χ^2 surface plotted versus a and b.

Table 7. Calculation of χ^2 for all three
roots of the system 22.

Parameter	set 1	set 2	set 3
a	-1.393	-1.463	-4.643
b	-0.861	0.321	3.539
χ^2	0.732	0.668	0.711

$$\begin{cases} \dfrac{64(2a+6b+1)(2a+2b+3)}{27(2a+3b+2)^2} = \dfrac{(K_1-\lambda T)K_3}{K_2^2} = \chi_3 \\[3mm] \dfrac{4(2a+6b+1)^2(4a+3b+8)}{(2a+3b+2)^3} = \dfrac{(K_1-\lambda T)^2 K_4}{K_2^3} = \chi_4 \end{cases} \quad (22)$$

Conclusions

1. In theory, if we have molecular system with n-components, IG can be obtained as solution of system of equations 15.

2. A non iterative, straightforward method of IG calculation for one and two-component system is proposed.

3. Applicability of method was verified with testing on simulated and measured data.

4. A rule of selecting of IG from possible results is suggested.

5. Impact of possible discontinuity points in model domain and consequently in χ^2 space is discussed

6. developed IG allows to increase the speed of analysis in most cases at least into 5 times.