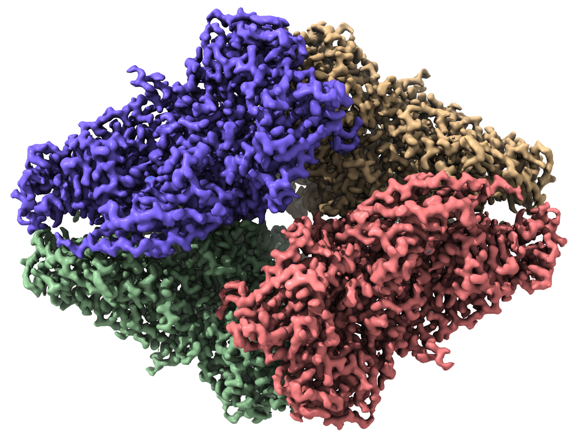
Inverse problems in Single Particle Analysis

FACULTY OF MATHEMATICS AND PHYSICS
Charles University

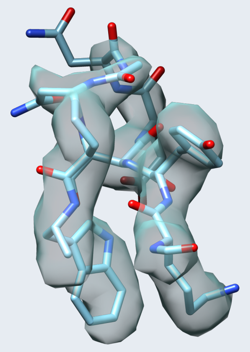


Eva Havelková and Iveta Hnětynková

Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University, Prague

Introduction

Cryo-electron microscopy Single Particle Analysis aims at the reconstruction of three-dimensional structure of biological macromolecules from a set of its two-dimensional projections created by a transmission electron microscope. Structural information is vital to understanding biological life processes and can be used, for example, in drug discovery. A wide variety of mathematical problems arises in the reconstruction process. These problems are challenging especially due to large dimensionality of the measured data, suffering from the presence of extreme amounts of noise and missing information. Here, we focus on an inverse problem of volume reconstruction from a given set of measured particle projections with estimated viewing angles.



Model formulation

The considered relation between the studied volume and its projection from a viewing angle ω is

$$g_\omega = PSF_\omega * E_\omega * (P_\omega f + e^s_\omega) + e^b_\omega, \qquad (1)$$

with

- PSF_ω, E_ω: Point Spread Functions.
- P_ω: X-ray transform: $P_\omega f(s) := \int_{-\infty}^{\infty} f(t\cdot\omega + s)\mathrm{d}t, \quad s \in \omega^\perp$.
- f: Unknown function representing the particle.
- g_ω: Measured projection (data).
- e^s_ω, e^b_ω: Structure and background noise functions.
- $*$: Convolution operator.

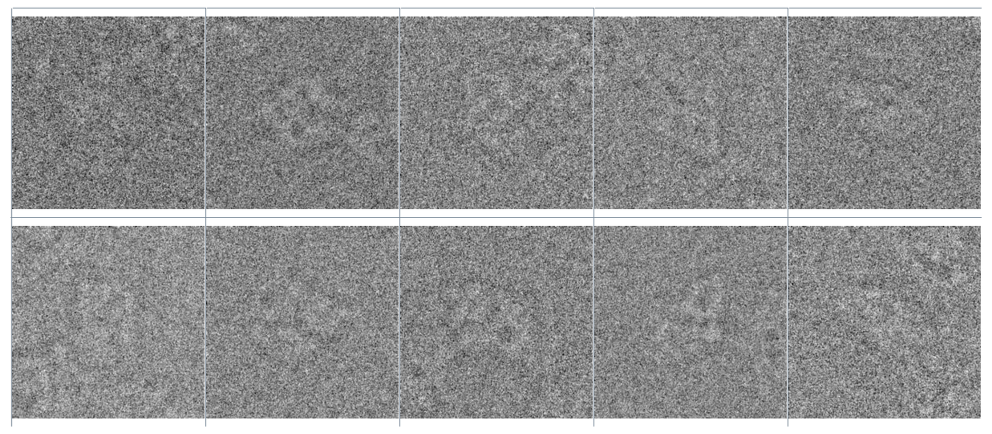
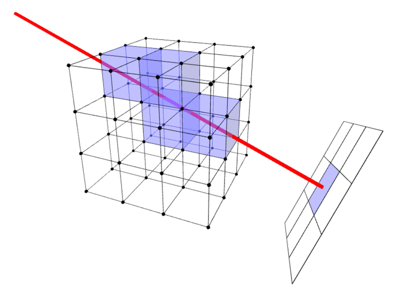


Figure: Input data - particle projections of the β-galactosidase macro-molecule measured by transmission electron microscope.

Discretization

In a discrete setting the aim is to reconstruct the values of the function f in a set of n^3 equidistantly distributed sampling points z_i. Typically, n is of order 10^2. Discretization of the model formulation (1) gives a system of linear equations



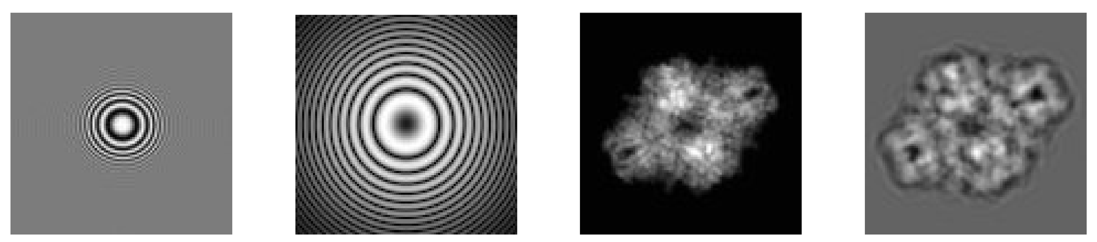
$$C_\omega P_\omega x = b_\omega, \qquad (2)$$

where x is a vector of the unknown values of f in the sampling points z_i and

- C_ω: BCCB matrix, analytic formula is known in the Fourier spectrum (CTF).
- P_ω: Sparse rectangular matrix containing the weights of a numerical quadrature

$$\int_{-\infty}^{\infty} f(t\cdot\omega + s)dt \approx \sum_{i=1}^{n^3} \alpha_i f(z_i).$$

- b_ω: Measured noisy projection.



(a) PSF (b) CTF (c) $P_\omega x$ (d) $C_\omega P_\omega x$

Figure: Example of PSF and its Fourier transform (CTF). $P_\omega x$ is a projection of β-gal. and $C_\omega P_\omega x$ is its blurred projection.

Discrete inverse problem

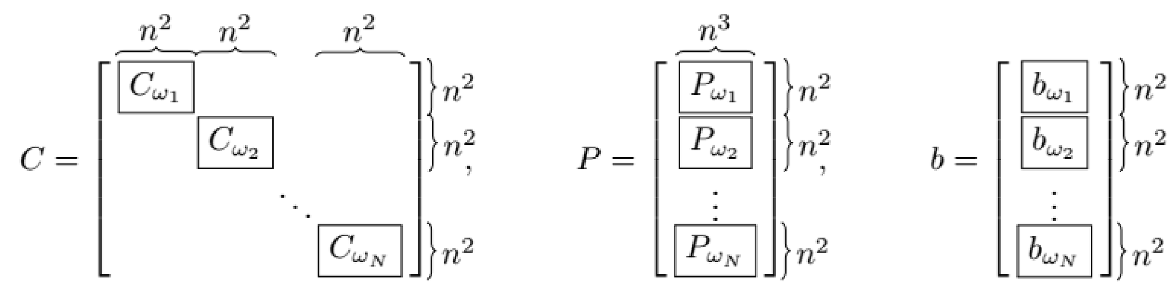
Considering N particle projections we obtain

$$CPx \approx b, \quad b = b^{exact} + e, \qquad (3)$$

where $C \in \mathbb{R}^{Nn^2 \times Nn^2}$, $P \in \mathbb{R}^{Nn^2 \times n^3}$, $x \in \mathbb{R}^{n^3}$, $b \in \mathbb{R}^{Nn^2}$. The vector e comprises noise in the data and it is assumed to be Gaussian white noise. The level of noise in the data is typically very high

$$\frac{\|e\|}{\|b^{exact}\|} > 1.$$

The number of particle projections N thus needs to be very high (order of 10^6) in order to obtain a good quality reconstruction.



$$C = \begin{bmatrix} C_{\omega_1} & & \\ & C_{\omega_2} & \\ & & \ddots \\ & & & C_{\omega_N} \end{bmatrix}, \quad P = \begin{bmatrix} P_{\omega_1} \\ P_{\omega_2} \\ \vdots \\ P_{\omega_N} \end{bmatrix}, \quad b = \begin{bmatrix} b_{\omega_1} \\ b_{\omega_2} \\ \vdots \\ b_{\omega_N} \end{bmatrix}$$

Figure: Illustration of the block structure of matrices C and P and right hand side vector b. C is a square matrix, P is overdetermined.

Iterative hybrid regularization

Hybrid LSQR method combining Krylov subspace method with Tikhonov regularization can be applied to solve the problem (3). Its key parts are:

- Golub-Kahan iterative bidiagonalization

$$L_{k+}y_k \approx \beta_1 e_1.$$

- Subsequent application of Tikhonov reg.

$$\min_y\{\|L_{k+}y - \beta_1 e_1\|_2^2 + \lambda_k^2\|y\|_2^2\}.$$

- Problem driven selection of λ_k.
  - Extracts solution norm estimate $\eta \approx \|x^{exact}\|_2$ from preceding reconstruction steps.
  - Selects λ_k such that $\|y_k\|_2 \approx \eta$.
- Stopping criterion (choice of the parameter k).
  - Monitors stagnation of $\|r_k\|_2$ - indicates stagnation of the method.

GPU implementation

Matrices C and P cannot be stored explicitly, their realistic size is hundreds of PB. Highly-parallel implementation of matrix-vector products utilizing GPUs and the block structure is thus necessary. Single precision arithmetic is typically used.

- Matrix-vector products with C and C^T:
  - Fast fourier transform and convolution theorem.
  - $O(n^2 log(n))$ operations per block.
- Matrix-vector products with P and P^T:
  - Fast voxel-traversal algorithms on GPU.
  - $P^T v$ is realized through row-wise access to P.
  - $O(n)$ operations per row.

Solution quality assesement

Evaluation of quality of the obtained solution is extremely challenging. Visualization tools are often used in practice.

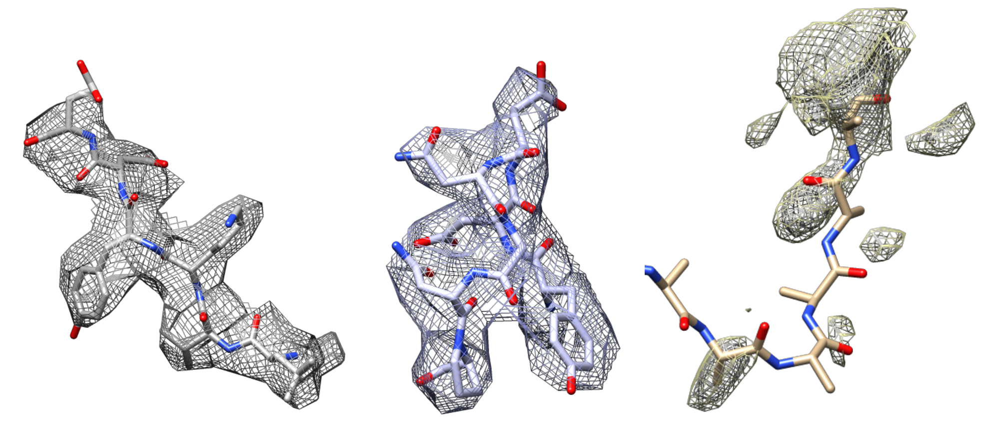


Figure: Selected parts of the computed reconstruction of β-gal.

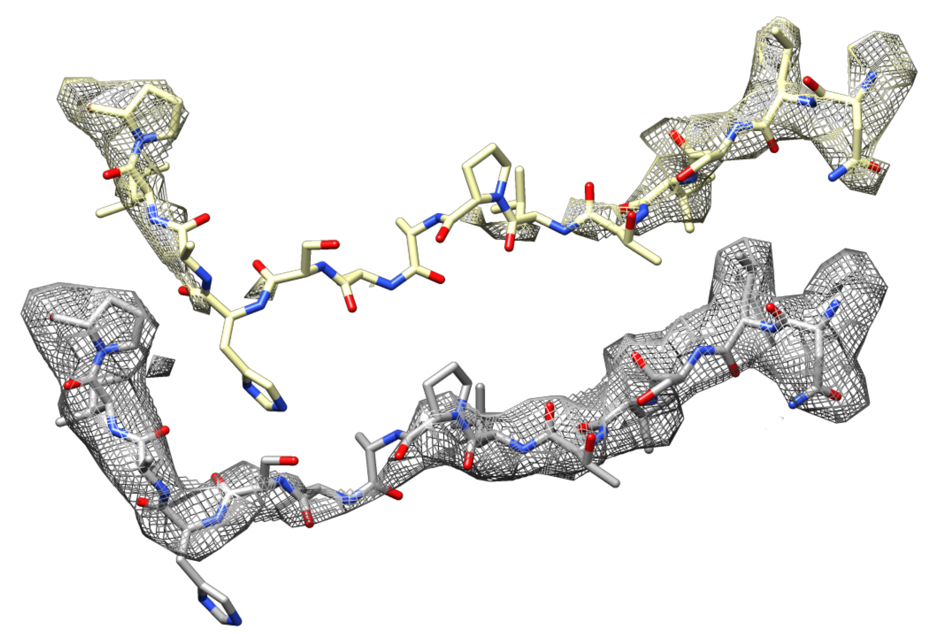


Figure: Comparison of selected parts computed by different methods.

References

[1] Bartesaghi A. Bendory T. and Singer A.
Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities.
IEEE Signal Processing Magazine, 37:58–76, 2020.

[2] E. Havelková and I. Hnětynková.
Iterative hybrid regularization for extremely noisy full models in single particle analysis.
Linear Algebra and its Applications (Accepted), 2022.

Acknowledgements

This work was partially supported by the Charles University, project GAUK 376121. The authors would also like to thank the Eyen SE company for providing the data, hardware and support.

Contact Information

- Email: eva.havelkova@karlin.mff.cuni.cz