Electron tomography: Mathematical challenges and approaches

Ozan Öktem
ozan.oktem@sidec.com
Sidec

January 20, 2009
ICES and CVC, University of Texas, Austin
Outline

1. The main application
2. The model for image formation and the reconstruction problem in ET
3. Reconstruction methods in ET
4. Statistical regularisation theory
The structure determination problem

Recover the 3D structure of an individual molecule (e.g. a protein or a macromolecular assembly) at highest possible resolution *in-situ* (in their cellular environment) or *in-vitro* (in aqueous environment).

- **X-ray crystallography** and **NMR** are established methods.
  1. Major advantage: Atomic resolution.
  2. Major disadvantage: Inability to study individual molecules in their natural environment (*in-situ* and *in-vitro*).

- **Electron tomography (ET)** is an emerging technology.
  1. Major advantage: Enables study of individual molecules in their natural environment (*in-situ* and *in-vitro*).
  2. Major disadvantage: Low resolution.
The structure determination problem

Recover the 3D structure of an individual molecule (e.g., a protein or a macromolecular assembly) at highest possible resolution \textit{in-situ} (in their cellular environment) or \textit{in-vitro} (in aqueous environment).

- X-ray crystallography and NMR are established methods.
  1. Major advantage: Atomic resolution.
  2. Major disadvantage: Inability to study individual molecules in their natural environment (\textit{in-situ} and \textit{in-vitro}).

- Electron tomography (ET) is an emerging technology.
  1. Major advantage: Enables study of individual molecules in their natural environment (\textit{in-situ} and \textit{in-vitro}).
  2. Major disadvantage: Low resolution.
The structure determination problem

Recover the 3D structure of an individual molecule (e.g. a protein or a macromolecular assembly) at highest possible resolution \textit{in-situ} (in their cellular environment) or \textit{in-vitro} (in aqueous environment).

- **X-ray crystallography** and NMR are established methods.
  1. Major advantage: Atomic resolution.
  2. Major disadvantage: Inability to study individual molecules in their natural environment (\textit{in-situ} and \textit{in-vitro}).

- **Electron tomography (ET)** is an emerging technology.
  1. Major advantage: Enables study of individual molecules in their natural environment (\textit{in-situ} and \textit{in-vitro}).
  2. Major disadvantage: Low resolution.
The forward model

Overview

The forward model can be divided into the following parts.

1. Electron-specimen interaction.
2. Optics of the TEM.
3. The intensity and detector response.
Electron-specimen interaction

Basic assumptions

- **One** electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^\text{in}$ is time harmonic and of the form

$$\psi^\text{in}(x, t) := e^{-ik^2 \frac{\hbar}{2m} t} u^\text{in}(x).$$
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coulomb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^\text{in}$ is time harmonic and of the form

$$\Psi^\text{in}(x, t) := e^{-ik^2\frac{\hbar}{2m} t} u^\text{in}(x).$$
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- **Coloumb potential** models elastic interaction, **absorption potential** models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^\text{in}$ is time harmonic and of the form

  $$\Psi^\text{in}(x, t) := e^{-ik^2 \frac{\hbar}{2m} t} u^\text{in}(x).$$
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^{\text{in}}$ is time harmonic and of the form

$$\Psi^{\text{in}}(x, t) := e^{-ik^2 \frac{\hbar}{2m} t} u^{\text{in}}(x).$$
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^\text{in}$ is time harmonic and of the form
  \[ \Psi^\text{in}(x, t) := e^{-i k^2 \frac{\hbar}{2m} t} u^\text{in}(x). \]
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^{\text{in}}$ is time harmonic and of the form

$$\Psi^{\text{in}}(x, t) := e^{-i k^{2} \frac{\hbar}{2 m} t} u^{\text{in}}(x).$$

Electron mass at rest
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^\text{in}$ is time harmonic and of the form

$$\Psi^\text{in}(x, t) := e^{-ik^2 \frac{h}{2m} t} u^\text{in}(x).$$

Amplitude term that fulfils the Helmholtz equation.
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^\text{in}$ is time harmonic and of the form

$$\Psi^\text{in}(x, t) := e^{-i \frac{k^2 \frac{\hbar}{2m}}{2} t} u^\text{in}(x).$$

- Perfect coherent illumination, i.e. $u^\text{in}(x) := e^{i k x \cdot \omega}$ where $\omega$ is the beam direction.
Electron-specimen interaction

Basic assumptions

- One electron in the specimen at a time. Average distance between two successive electrons much greater than the specimen thickness.
- Interaction is governed by the scalar Schrödinger wave equation.
- Coloumb potential models elastic interaction, absorption potential models decrease in the flux of the non-scattered and elastically scattered electrons.
- Incident wave $\Psi^{\text{in}}$ is time harmonic and of the form
  \[ \Psi^{\text{in}}(x, t) := e^{-i k^2 \frac{\hbar}{2m} t} u^{\text{in}}(x). \]

- Perfect coherent illumination, i.e. $u^{\text{in}}(x) := e^{i k x \cdot \omega}$ where $\omega$ is the beam direction.
- Specimen is weakly scattering so first order Born approximation is valid and we can linearise the intensity.
Optics of the TEM

The optical setup and assumptions

System is aligned w.r.t. optical axis $\omega$
Optics of the TEM

The optical setup and assumptions

Ideal thin lens. Focal length $f$ and spherical aberration $C_s$ equals that of real objective lens, magnification equals that of the entire microscope.
Optics of the TEM

The optical setup and assumptions

System is aligned w.r.t. optical axis

Ideal thin lens. Focal length $f$ and spherical aberration $C_s$ equals that of real objective lens, magnification equals that of the entire microscope.

Lens

Object plane

Image plane

System is in focus, i.e. \[ \frac{1}{r} + \frac{1}{q} = \frac{1}{f} \]
Optics of the TEM

The optical setup and assumptions

System is aligned w.r.t. optical axis

\[ \omega \perp - q\omega \quad \omega \perp \quad \omega \perp + f\omega \]

Object plane

Focal plane

Aperture (rotation invariant)
Optics of the TEM
The optical setup and assumptions

System is aligned w.r.t. optical axis

Ideal thin lens. Focal length $f$ and spherical aberration $C_s$ equals that of real objective lens, magnification equals that of the entire microscope.

Lens

Object plane $\omega^\perp - q\omega$  Lens  Focal plane $\omega^\perp + f\omega$  Image plane $\omega^\perp + r\omega$
Optics of the TEM
The optical setup and assumptions

\[
\omega^\perp - q\omega \quad \omega^\perp \quad \omega^\perp + f\omega \quad \omega^\perp + r\omega
\]

Object plane  \quad Lens  \quad Focal plane  \quad Image plane

System is in focus, i.e.  \( \frac{1}{r} + \frac{1}{q} = \frac{1}{f} \)
The intensity

The standard model for image formation

Standard model for image formation in ET is based on taking the first term in the asymptotic expansion of the propagation operator as $k \to \infty$.

\[
I(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right) \left[ F(x) := -\frac{2m}{\hbar^2} (V(x) + i\Lambda(x)) \right]
\]

Coloumb potential

Absorption potential
The intensity

The standard model for image formation

Standard model for image formation in ET is based on taking the first term in the asymptotic expansion of the propagation operator as $k \to \infty$.

The unscattered wave

$$I(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right)$$

The intensity

The magnification
The intensity

The standard model for image formation

**Standard model for image formation in ET** is based on taking the first term in the asymptotic expansion of the propagation operator as $k \to \infty$.

$$I(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right) \left\{ \text{PSF}^\text{re}_k(\omega, \cdot) \otimes \mathcal{P}(F^\text{re})(\omega, -\cdot) \right\} \left( \frac{z}{M} \right) +$$

where $\mathcal{P}$ denotes the X-ray transform.
The intensity

The standard model for image formation

Standard model for image formation in ET is based on taking the first term in the asymptotic expansion of the propagation operator as \( k \to \infty \).

\[
\mathcal{I}(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right) \left\{ \text{PSF}^\text{re}_k(\omega, \cdot) \odot \mathcal{P}(F^\text{re})(\omega, -\cdot) \right\} \left( \frac{z}{M} \right) +
\]

The characteristic function for the aperture

where \( \mathcal{P} \) denotes the X-ray transform and

\[
\text{PSF}^\text{re}_k(\omega, y) := \mathcal{F}_{\omega^\perp} \left\{ \chi \left( \frac{f}{k} \cdot + f\omega \right) \sin \left[ \gamma_k \left( | \cdot |^2 \right) \right] \right\}(y)
\]
The intensity

The standard model for image formation

Standard model for image formation in ET is based on taking the first term in the asymptotic expansion of the propagation operator as $k \to \infty$.

$$ I(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right) \left\{ \text{PSF}^\text{re}_k(\omega, \cdot) \ast \mathcal{P}(F^\text{re})(\omega, - \cdot) \right\} \left( \frac{z}{M} \right) + $$

$$ \gamma_k(t) := -t \frac{1}{4k} \left( C_s k^{-2} t - 2\Delta z \right) $$

where $\mathcal{P}$ denotes the X-ray transform and

$$ \text{PSF}^\text{re}_k(\omega, y) := \mathcal{F}_{\omega} \left\{ \chi \left( \frac{f}{k} \cdot + f\omega \right) \sin \left[ \gamma_k (|\cdot|^2) \right] \right\}(y) $$
The intensity

The standard model for image formation

**Standard model for image formation in ET** is based on taking the first term in the asymptotic expansion of the propagation operator as \( k \to \infty \).

\[
\mathcal{I}(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right) \left\{ \text{PSF}_{k}^{\text{re}}(\omega, \cdot) \otimes \mathcal{P}(F^{\text{re}}(\omega, -\cdot)) \right\} \left( \frac{z}{M} \right) +
\]

\[
\gamma_k(t) := -t \frac{1}{4k} \left( C_s k^{-2} t - 2 \Delta z \right)
\]

Spherical aberration

Defocus

where \( \mathcal{P} \) denotes the X-ray transform and

\[
\text{PSF}_{k}^{\text{re}}(\omega, y) := \mathcal{F}_{\omega \perp} \left\{ \chi \left( \frac{f}{k} \cdot + f\omega \right) \sin \left[ \gamma_k \left( | \cdot |^2 \right) \right] \right\}(y)
\]
The intensity

The standard model for image formation

Standard model for image formation in ET is based on taking the first term in the asymptotic expansion of the propagation operator as $k \to \infty$.

$$I(F)(\omega, z) = \frac{1}{M^2} \left( 1 - 2(2\pi)^{-2} \right) \left[ \left. \begin{array}{c} \text{PSF}^{\text{re}}_k(\omega, \cdot) \ast \mathcal{P}(F^{\text{re}})(\omega, - \cdot) \right|_{\omega_{\perp}} \left( \frac{z}{M} \right) + \\ \left. \text{PSF}^{\text{im}}_k(\omega, \cdot) \ast \mathcal{P}(F^{\text{im}})(\omega, - \cdot) \right|_{\omega_{\perp}} \left( \frac{z}{M} \right) \right] k^{-1}$$

where $\mathcal{P}$ denotes the X-ray transform and

$$\text{PSF}^{\text{im}}_k(\omega, y) := \mathcal{F}_{\omega_{\perp}} \left\{ \chi \left( \frac{f}{k} \cdot + f\omega \right) \cos \left[ \gamma_k (| \cdot |^2) \right] \right\}(y)$$
The forward operator and the measured data

A single measured data point is a sample from the random variable

\[ \text{data}[F](\omega, z) := \left\{ \mathcal{F}_{\omega}^{-1}[\text{MTF}](\cdot) \ast c[F](\omega, \cdot) \right\}(z) + E(\omega, z) \]
The forward operator and the measured data

A single measured data point is a sample from the random variable

\[ \text{data}[F](\omega, z) := \left\{ \mathcal{F}_{\omega}^{-1}[\text{MTF}](\cdot) \ast \omega_{\perp} c[F](\omega, \cdot) \right\}(z) + E(\omega, z) \]

\( c[F](\omega, z) \) is a random variable representing the stochasticity in the counting process. It has distribution

\[ c[F](\omega, z) \sim \text{Poisson}\left[ \text{Dose}(\omega) \mathcal{I}(F)(\omega, z) \right] \]

with \( \text{Dose}(\omega) \) representing the incoming dose of electrons per pixel for the direction \( \omega \).
The forward operator and the measured data

A single measured data point is a sample from the random variable

\[ \text{data}[F](\omega, z) := \left\{ \mathcal{F}\omega \perp [\text{MTF}](\cdot) \otimes c[F](\omega, \cdot) \right\}(z) + E(\omega, z) \]

\( E(\omega, z) \) is a random variable that represents the additive measurement error from the detector. It is usually Gaussian or uniformly distributed and precise distribution is estimated from specific tests made on the detector.
The forward operator and the measured data

A single measured data point is a sample from the random variable

\[
\text{data}[F](\omega, z) := \left\{ \mathcal{F}_{\omega \perp}^{-1}[\text{MTF}](\cdot) \otimes c[F](\omega, \cdot) \right\}(z) + E(\omega, z)
\]

MTF is the modulation transfer function which essentially describes how the CCD camera attenuates different spatial frequencies present in the input signal.
The forward operator and the measured data

A single measured data point is a sample from the random variable

$$data[F](\omega, z) := \left\{ \mathcal{F}_{\omega \perp}^{-1}[MTF](\cdot) \ast \omega \perp \right\} \ast c[F](\omega, \cdot) (z) + E(\omega, z)$$

The measured data from an ET experiment is modelled as a sample of the \(mn^2\)-dimensional random variable \(data[F]\) defined as

$$data[F] := \left( data[F](\omega_1, z_{1,1}), \ldots, data[F](\omega_m, z_{m,n^2}) \right).$$

where \(\{\omega_1, \ldots, \omega_m\} \subset S^2\) are the different directions and for each direction \(\omega_j\) we have \(n^2\) pixels with midpoints \(\{z_{j,1}, \ldots, z_{j,n^2}\} \subset \omega_j \perp\).
The forward operator and the measured data

A single measured data point is a sample from the random variable

$$\text{data}[F](\omega, z) := \left\{ \mathcal{F}_{\omega_j}^{-1}[\text{MTF}](\cdot) \ast_{\omega_j} c[F](\omega, \cdot) \right\}(z) + E(\omega, z)$$

The measured data from an ET experiment is modelled as a sample of the $mn^2$-dimensional random variable $\text{data}[F]$ defined as

$$\text{data}[F] := \left( \text{data}[F](\omega_1, z_{1,1}), \ldots, \text{data}[F](\omega_m, z_{m,n^2}) \right).$$

where $\{\omega_1, \ldots, \omega_m\} \subset S^2$ are the different directions and for each direction $\omega_j$ we have $n^2$ pixels with midpoints $\{z_{j,1}, \ldots, z_{j,n^2}\} \subset \omega_j^\perp$. The forward operator $\mathcal{T}(F)(\omega, z)$ is given as the expected value of $\text{data}[F](\omega, z)$ and the forward data operator is defined as

$$\mathbf{T}(F) := \left( \mathcal{T}(F)(\omega_1, z_{1,1}), \ldots, \mathcal{T}(F)(\omega_m, z_{m,n^2}) \right).$$
The reconstruction problem in ET
Formulation and difficulties

The reconstruction problem in ET

Recover the scattering potential $f$ from the measured tiltseries $g$ which is a single sample of the stochastic variable $G$ modelling the data.

Any attempt at solving the reconstruction problem in ET has to deal with the following:

The dose problem: Total dose is about 500–1250 e$^-$/pixel (at 25000× magnification) distributed over 60 or 120 tilts, so one has data with significant amount of Poisson stochasticity.

Limited angle problem: Leads to severe instability.

Region of interest problem: Leads to non-uniqueness.

The combination of data with very low signal-to-noise (dose problem) and severe ill-posedness (limited angle and region of interest problems) in the reconstruction problem is devastating!
The reconstruction problem in ET

Formulation and difficulties

Recover the scattering potential $f$ from the measured tiltseries $g$ which is a single sample of the stochastic variable $G$ modelling the data.

Any attempt at solving the reconstruction problem in ET has to deal with the following:

The dose problem: Total dose is about 500–1250 e$^-$/pixel (at 25000× magnification) distributed over 60 or 120 tilts, so one has data with significant amount of Poisson stochasticity.

Limited angle problem: Leads to severe instability.

Region of interest problem: Leads to non-uniqueness.

The combination of data with very low signal-to-noise (dose problem) and severe ill-posedness (limited angle and region of interest problems) in the reconstruction problem is devastating!
The reconstruction problem in ET

Formulation and difficulties

The reconstruction problem in ET

Recover the scattering potential $f$ from the measured tiltseries $g$ which is a single sample of the stochastic variable $G$ modelling the data.

Any attempt at solving the reconstruction problem in ET has to deal with the following:

**The dose problem:** Total dose is about $500–1250 \text{ e}^{-/\text{pixel}}$ (at $25000 \times \text{magnification}$) distributed over 60 or 120 tilts, so one has data with significant amount of Poisson stochasticity.

**Limited angle problem:** Leads to severe instability.

**Region of interest problem:** Leads to non-uniqueness.

The combination of data with very low signal-to-noise (dose problem) and severe ill-posedness (limited angle and region of interest problems) in the reconstruction problem is devastating!
The reconstruction problem in ET
Formulation and difficulties

Recover the scattering potential $f$ from the measured tiltseries $g$ which is a single sample of the stochastic variable $G$ modelling the data.

Any attempt at solving the reconstruction problem in ET has to deal with the following:

The dose problem: Total dose is about $500–1250 \text{ e}^-/\text{pixel}$ (at $25000\times$ magnification) distributed over 60 or 120 tilts, so one has data with significant amount of Poisson stochasticity.

Limited angle problem: Leads to severe instability.

Region of interest problem: Leads to non-uniqueness.

The combination of data with very low signal-to-noise (dose problem) and severe ill-posedness (limited angle and region of interest problems) in the reconstruction problem is devastating!
The reconstruction problem in ET

Example of a TEM image

Figure: TEM (200 kV) images of an *in-vitro* sample containing microsomes (small spherical shells with a diameter of about 60 nm). Left image is a low dose image (about 3000 e⁻/nm²) taken at 1 µm under-focus used for reconstruction, right image is a high dose image of same area taken at 12 µm under-focus. Length of sides are 150 nm ×106 nm (H×W).
The reconstruction problem in ET

Further difficulties

Besides the severe ill-posedness, there are two additional peculiarities in ET that makes it difficult to apply an off-the-shelf regularisation method.

- **No reliable a priori estimate of data error.** This makes it difficult to *a priori* determine the degree of stabilisation (*i.e.* the regularisation parameter(s)).

- In ET we are dealing with a multi-component reconstruction problem, *i.e.* there are parameters which are additional unknowns that needs to be reconstructed alongside the scattering potential:
  - Independent of specimen, dependent on tilt: TEM imaging parameters, *e.g.* incoming dose and defocus.
  - Dependent of specimen: Parameters used for describing the scattering potential outside the region of interest and the (tilt dependent) phase contrast ratio.

Note: Parameters independent of the specimen and tilt, *e.g.* detector related parameters, can be determined beforehand.
The reconstruction problem in ET

Further difficulties

Besides the severe ill-posedness, there are two additional peculiarities in ET that makes it difficult to apply an off-the-shelf regularisation method.

- No reliable *a priori* estimate of data error. This makes it difficult to *a priori* determine the degree of stabilisation (*i.e.* the regularisation parameter(s)).
- In ET we are dealing with a multi-component reconstruction problem, *i.e.* there are parameters which are additional unknowns that needs to be reconstructed alongside the scattering potential:
  - Independent of specimen, dependent on tilt: TEM imaging parameters, *e.g.* incoming dose and defocus.
  - Dependent of specimen: Parameters used for describing the scattering potential outside the region of interest and the (tilt dependent) phase contrast ratio.

Note: Parameters independent of the specimen and tilt, *e.g.* detector related parameters, can be determined beforehand.
Reconstruction methods in ET

Overview of approaches

\(\mathcal{T}\) denotes the forward operator which is defined as the expected value of the stochastic variable \(G\) modelling the data (i.e. the measured tiltseries \(g\) is a sample of \(G\)).

Analytic methods: Based on discretising the inverse of \(\mathcal{T}\). Regularisation is obtained by constructing an approximate inverse recovering only those features that can be stably retrieved.
Example: Filter backprojection (FBP) with de-noising, \(\Lambda\)-tomography for the recovery of boundaries (edges).

Iterative methods: Discretise \(\mathcal{T}(f) = g\) and solve by an iterative method. Regularisation is obtained by early stopping.
Example: ART, SIRT

Variational regularisation methods
Reconstruction methods in ET

Overview of approaches

\( T \) denotes the forward operator which is defined as the expected value of the stochastic variable \( G \) modelling the data \((i.e. \text{the measured tiltseries } g \text{ is a sample of } G)\).

**Analytic methods:** Based on discretising the inverse of \( T \). Regularisation is obtained by constructing an approximate inverse recovering only those features that can be stably retrieved. **Example:** Filter backprojection (FBP) with de-noising, \( \Lambda \)-tomography for the recovery of boundaries (edges).

**Iterative methods:** Discretise \( T(f) = g \) and solve by an iterative method. Regularisation is obtained by early stopping. **Example:** ART, SIRT

**Variational regularisation methods**
Reconstruction methods in ET

Overview of approaches

\( \mathcal{T} \) denotes the forward operator which is defined as the expected value of the stochastic variable \( G \) modelling the data (i.e. the measured tiltseries \( g \) is a sample of \( G \)).

**Analytic methods:** Based on discretising the inverse of \( \mathcal{T} \). Regularisation is obtained by constructing an approximate inverse recovering only those features that can be stably retrieved.

**Example:** Filter backprojection (FBP) with de-noising, \( \Lambda \)-tomography for the recovery of boundaries (edges).

**Iterative methods:** Discretise \( \mathcal{T}(f) = g \) and solve by an iterative method. Regularisation is obtained by early stopping.

**Example:** ART, SIRT

**Variational regularisation methods**
Reconstruction methods in ET
Overview of approaches

\( \mathcal{T} \) denotes the forward operator which is defined as the expected value of the stochastic variable \( G \) modelling the data (i.e. the measured tiltseries \( g \) is a sample of \( G \)).

**Analytic methods**: Based on discretising the inverse of \( \mathcal{T} \). Regularisation is obtained by constructing an approximate inverse recovering only those features that can be stably retrieved.

**Example**: Filter backprojection (FBP) with de-noising, \( \Lambda \)-tomography for the recovery of boundaries (edges).

**Iterative methods**: Discretise \( \mathcal{T}(f) = g \) and solve by an iterative method. Regularisation is obtained by early stopping.

**Example**: ART, SIRT

Variational regularisation methods
Classical regularisation theory

Abstract reconstruction problem

The abstract reconstruction problem

Assume that \( g \in \mathcal{H} \) is related to \( f \in \mathcal{X} \) by the model \( T(f) = g \). Recover \( f \in \mathcal{X} \) from \( g \in \mathcal{H} \).

In the above:

- \( \mathcal{H} \) is the data space, i.e. the set of all possible data, and \( g \in \mathcal{H} \) is the measured data. We will henceforth assume that \( \mathcal{H} \cong \mathbb{R}^m \) where \( m \) is the number of data points.
- \( \mathcal{X} \) is the reconstruction space, i.e. the set of feasible solutions, and \( f \in \mathcal{X} \) represents the unknown object of primary interest that is to be recovered. In ET, \( \mathcal{X} \) are functions in \( L^1 \cap L^2 \) with positive real and imaginary parts.
- \( T : \mathcal{X} \rightarrow \mathcal{H} \) is the forward operator which is the mathematical model for the experiment relating the unknown \( f \) to the measured data \( g \).
Classical regularisation theory

Abstract reconstruction problem

The abstract reconstruction problem

Assume that $g \in \mathcal{H}$ is related to $f \in \mathcal{X}$ by the model $T(f) = g$. Recover $f \in \mathcal{X}$ from $g \in \mathcal{H}$.

In the above:

- $\mathcal{H}$ is the data space, i.e. the set of all possible data, and $g \in \mathcal{H}$ is the measured data. We will henceforth assume that $\mathcal{H} \simeq \mathbb{R}^m$ where $m$ is the number of data points.

- $\mathcal{X}$ is the reconstruction space, i.e. the set of feasible solutions, and $f \in \mathcal{X}$ represents the unknown object of primary interest that is to be recovered. In ET, $\mathcal{X}$ are functions in $L^1 \cap L^2$ with positive real and imaginary parts.

- $T : \mathcal{X} \rightarrow \mathcal{H}$ is the forward operator which is the mathematical model for the experiment relating the unknown $f$ to the measured data $g$. 
Classical regularisation theory

Abstract reconstruction problem

The abstract reconstruction problem

Assume that \( g \in \mathcal{H} \) is related to \( f \in \mathcal{X} \) by the model \( T(f) = g \). Recover \( f \in \mathcal{X} \) from \( g \in \mathcal{H} \).

In the above:

- \( \mathcal{H} \) is the data space, i.e. the set of all possible data, and \( g \in \mathcal{H} \) is the measured data. We will henceforth assume that \( \mathcal{H} \cong \mathbb{R}^m \) where \( m \) is the number of data points.

- \( \mathcal{X} \) is the reconstruction space, i.e. the set of feasible solutions, and \( f \in \mathcal{X} \) represents the unknown object of primary interest that is to be recovered. In ET, \( \mathcal{X} \) are functions in \( L^1 \cap L^2 \) with positive real and imaginary parts.

- \( T : \mathcal{X} \to \mathcal{H} \) is the forward operator which is the mathematical model for the experiment relating the unknown \( f \) to the measured data \( g \)
Classical regularisation theory

Concept of ill-posedness

An important problem that arises when trying to solve

\[ \mathcal{I}(f) = g \]  \hspace{1cm} (1)

is that there might be no solutions at all (non-existence).

Existence is commonly enforced by considering least squares solutions to (1), i.e. (1) is replaced with

\[ \min_{f \in \mathcal{X}} \| \mathcal{I}(f) - g \|_{\mathcal{H}} \]

for a suitable choice of distance measure \( \| \cdot \|_{\mathcal{H}} \) in data space \( \mathcal{H} \).

Two serious problems remain:

- Infinitely many least squares solutions (non-uniqueness).
- The least squares solution does not depend continuously on the data \( g \) (instability).

(1) is ill-posed if any of these issues occur.
Classical regularisation theory

Concept of ill-posedness

An important problem that arises when trying to solve

\[ T(f) = g \]  \hspace{1cm} (1)

is that there might be no solutions at all (non-existence). Existence is commonly enforced by considering least squares solutions to \( (1) \), i.e. \( (1) \) is replaced with

\[
\min_{f \in X} \| T(f) - g \|_H
\]

for a suitable choice of distance measure \( \| \cdot \|_H \) in data space \( H \).

Two serious problems remain:

- Infinitely many least squares solutions (non-uniqueness).
- The least squares solution does not depend continuously on the data \( g \) (instability).

\( (1) \) is ill-posed if any of these issues occur.
Classical regularisation theory

Concept of ill-posedness

An important problem that arises when trying to solve

\[ T(f) = g \]  \hspace{1cm} (1)  

is that there might be no solutions at all (non-existence). Existence is commonly enforced by considering least squares solutions to (1), i.e. (1) is replaced with

\[ \min_{f \in X} \| T(f) - g \|_H \]

for a suitable choice of distance measure \( \| \cdot \|_H \) in data space \( H \).

Two serious problems remain:

- Infinitely many least squares solutions (non-uniqueness).
- The least squares solution does not depend continuously on the data \( g \) (instability).

(1) is ill-posed if any of these issues occur.
Classical regularisation theory

Concept of ill-posedness

An important problem that arises when trying to solve

\[ T(f) = g \] (1)

is that there might be no solutions at all (non-existence). Existence is commonly enforced by considering least squares solutions to (1), i.e. (1) is replaced with

\[ \min_{f \in \mathcal{X}} \| T(f) - g \|_{\mathcal{H}} \]

for a suitable choice of distance measure \( \| \cdot \|_{\mathcal{H}} \) in data space \( \mathcal{H} \).

Two serious problems remain:

- Infinitely many least squares solutions (non-uniqueness).
- The least squares solution does not depend continuously on the data \( g \) (instability).

(1) is ill-posed if any of these issues occur.
Classical regularisation theory

Basic principle of regularisation

Special care must be taken when dealing with ill-posed reconstruction problems. The reconstruction method must involve some stabilisation, i.e. it must act as a regularisation method.

The main idea

The main idea underlying a regularisation method is to replace the original ill-posed reconstruction problem by a well-posed reconstruction problem (i.e. it has a unique solution that depends continuously on the data) that is convergent as the data error goes to zero.

Well-posedness: Guarantees stability

Convergence: The reconstructions obtained converge to a least squares solution when the data error approaches zero and the parameters in the reconstruction method (regularisation parameters) are chosen appropriately.
**Classical regularisation theory**

Basic principle of regularisation

Special care must be taken when dealing with ill-posed reconstruction problems. The reconstruction method must involve some stabilisation, *i.e.* it must act as a regularisation method.

---

**The main idea**

The main idea underlying a regularisation method is to replace the original ill-posed reconstruction problem by a well-posed reconstruction problem (*i.e.* it has a unique solution that depends continuously on the data) that is convergent as the data error goes to zero.

---

**Well-posedness:** Guarantees stability

**Convergence:** The reconstructions obtained converge to a least squares solution when the data error approaches zero and the parameters in the reconstruction method (regularisation parameters) are chosen appropriately.
Classical regularisation theory

Basic principle of regularisation

Special care must be taken when dealing with ill-posed reconstruction problems. The reconstruction method must involve some stabilisation, i.e. it must act as a regularisation method.

The main idea

The main idea underlying a regularisation method is to replace the original ill-posed reconstruction problem by a well-posed reconstruction problem (i.e. it has a unique solution that depends continuously on the data) that is convergent as the data error goes to zero.

Well-posedness: Guarantees stability

Convergence: The reconstructions obtained converge to a least squares solution when the data error approaches zero and the parameters in the reconstruction method (regularisation parameters) are chosen appropriately.
Besides convergence, the performance of a regularisation method depends on

- the accuracy of the forward operator $T$ modelling the experiment,
- how well it can account for the actual stochasticity of the data $g$,
- how well it captures a priori knowledge about the solution $f$, and
- on finding the right compromise between accuracy and stability by selecting the regularisation parameter(s).
Analytic methods: Filtered backprojection method

Basic assumptions

- Interpret data $g$ (possibly after post-processing) as projections of the scattering potential $f$, i.e. $g = \mathcal{P}(f)$.

- The reconstruction problem is to recover the scattering potential $f$ from the projections $g = \mathcal{P}(f)$.

- In ET, projections are sampled on a parallel beam line complex specified by a curve $S \subset S^2$ which is defined as the set of parallel lines where the direction vector is restricted to the curve $S$. 
Analytic methods: Filtered backprojection method

Basic assumptions

- Interpret data $g$ (possibly after post-processing) as projections of the scattering potential $f$, i.e., $g = \mathcal{P}(f)$.
- The reconstruction problem is to recover the scattering potential $f$ from the projections $g = \mathcal{P}(f)$.
- In ET, projections are sampled on a parallel beam line complex specified by a curve $S \subset S^2$ which is defined as the set of parallel lines where the direction vector is restricted to the curve $S$. 
Analytic methods: Filtered backprojection method

Basic assumptions

- Interpret data $g$ (possibly after post-processing) as projections of the scattering potential $f$, i.e. $g = \mathcal{P}(f)$.
- The reconstruction problem is to recover the scattering potential $f$ from the projections $g = \mathcal{P}(f)$.
- In ET, projections are sampled on a parallel beam line complex specified by a curve $S \subset S^2$ which is defined as the set of parallel lines where the direction vector is restricted to the curve $S$. 
Analytic methods: Filtered backprojection method

Reconstruction scheme

**Reconstruction operator:** For a function \( g \) defined on a parallel beam line complex specified by the curve \( S \subset S^2 \), define

\[
\mathcal{L}_{\text{FBP}}(g) := P^*_S(h \ast \omega g)
\]

where

- \( h \) is the reconstruction filter
- \( P^*_S \) is the back-projection restricted to the curve of directions \( S \):  

\[
P^*_S(g)(x) := \int_S g(\omega, x - (x \cdot \omega)\omega) \, d\omega \quad \text{where} \; x \in \mathbb{R}^3.
\]
Analytic methods: Filtered backprojection method

Reconstruction scheme

Reconstruction operator: For a function $g$ defined on a parallel beam line complex specified by the curve $S \subset S^2$, define

$$\mathcal{L}_{\text{FBP}}(g) := \mathcal{P}_S^*(h \ast g)$$

where

- $h$ is the reconstruction filter
- $\mathcal{P}_S^*$ is the back-projection restricted to the curve of directions $S$:

$$\mathcal{P}_S^*(g)(x) := \int_S g(\omega, x - (x \cdot \omega)\omega) \, d\omega \quad \text{where} \ x \in \mathbb{R}^3.$$  

Note: The pair $(\omega, x)$, where $\omega \in S$ is the direction of the line and $x \in \omega^\perp$ as a point on the line, uniquely determines a line and thereby acts as coordinates on the parallel beam line complex.
Analytic methods: Filtered backprojection method

Reconstruction scheme

Reconstruction operator: For a function $g$ defined on a parallel beam line complex specified by the curve $S \subset S^2$, define

$$\mathcal{L}_{\text{FBP}}(g) := \mathcal{P}_S^*(h \ast \perp g)$$

where

- $h$ is the reconstruction filter
- $\mathcal{P}_S^*$ is the back-projection restricted to the curve of directions $S$:

$$\mathcal{P}_S^*(g)(\mathbf{x}) := \int_S g(\omega, \mathbf{x} - (\mathbf{x} \cdot \omega) \omega) \, d\omega \quad \text{where} \quad \mathbf{x} \in \mathbb{R}^3.$$

Rationale: If $g = \mathcal{P}(f)$, then one can show that

$$\mathcal{L}_{\text{FBP}}(g) = H \ast f \quad \text{where} \quad H := \mathcal{P}_S^*(h).$$

Exact FBP: Choose $h$ such that $H = \delta$. 
Analytic methods: Filtered backprojection method

Problems

**Limited data:** Due to the limited angle problems in ET, $S$ does not fulfil Orlov’s condition and the reconstruction problem is severely ill-posed. Local tomography introduces non-uniqueness even for noise free continuous data.

**Difficult to regularise:** If data $g$ is not consistent, then discretised exact FBP gives a least squares solution. Thus, one must regularise. Crowther’s regularisation criterion based on band-limiting the filter doesn’t work well with very noisy data.

**Difficult to account for data stochasticity:** Difficult (if not impossible) to devise schemes for selecting the filter $h$ that takes into account the specific stochasticity of the data.

**Simplified model for imaging:** Data is assumed to be samples of a projection, so optics and detector PSF’s must either be deconvolved or ignored.
Analytic methods: Filtered backprojection method

Problems

**Limited data:** Due to the limited angle problems in ET, $S$ does not fulfill Orlov’s condition and the reconstruction problem is severely ill-posed. Local tomography introduces non-uniqueness even for noise free continuous data.

**Difficult to regularise:** If data $g$ is not consistent, then discretised exact FBP gives a least squares solution. Thus, one must regularise. Crowther’s regularisation criterion based on band-limiting the filter doesn’t work well with very noisy data.

**Difficult to account for data stochasticity:** Difficult (if not impossible) to devise schemes for selecting the filter $h$ that takes into account the specific stochasticity of the data.

**Simplified model for imaging:** Data is assumed to be samples of a projection, so optics and detector PSF’s must either be deconvolved or ignored.
Analytic methods: Filtered backprojection method

Problems

Limited data: Due to the limited angle problems in ET, $S$ does not fulfil Orlov’s condition and the reconstruction problem is severely ill-posed. Local tomography introduces non-uniqueness even for noise free continuous data.

Difficult to regularise: If data $g$ is not consistent, then discretised exact FBP gives a least squares solution. Thus, one must regularise. Crowther’s regularisation criterion based on band-limiting the filter doesn’t work well with very noisy data.

Difficult to account for data stochasticity: Difficult (if not impossible) to devise schemes for selecting the filter $h$ that takes into account the specific stochasticity of the data.

Simplified model for imaging: Data is assumed to be samples of a projection, so optics and detector PSF’s must either be deconvolved or ignored.
Analytic methods: Filtered backprojection method

Problems

**Limited data:** Due to the limited angle problems in ET, $S$ does not fulfil Orlov’s condition and the reconstruction problem is severely ill-posed. Local tomography introduces non-uniqueness even for noise free continuous data.

**Difficult to regularise:** If data $g$ is not consistent, then discretised exact FBP gives a least squares solution. Thus, one must regularise. Crowther’s regularisation criterion based on band-limiting the filter doesn’t work well with very noisy data.

**Difficult to account for data stochasticity:** Difficult (if not impossible) to devise schemes for selecting the filter $h$ that takes into account the specific stochasticity of the data.

**Simplified model for imaging:** Data is assumed to be samples of a projection, so optics and detector PSF’s must either be deconvolved or ignored.
Analytic methods: Λ-tomography

Basic assumptions and principle

In ET one has limited angle local tomography data. Therefore, one cannot exactly reconstruct the scattering potential of the specimen even in cases when one assumes to have a continuum of exact data!

Solution proposed by Λ-tomography

Reconstruct only some information about the specimen that can be stably retrieved, in our case the singularities of the scattering potential, i.e. the boundaries of the molecules in the specimen.

Λ-tomography is based on the same assumptions as FBP.

Reference(s)

Local Tomography in Electron Microscopy
Quinto E T and Öktem O
Analytic methods: Λ-tomography

Basic assumptions and principle

In ET one has limited angle local tomography data. Therefore, one cannot exactly reconstruct the scattering potential of the specimen even in cases when one assumes to have a continuum of exact data!

Solution proposed by Λ-tomography

Reconstruct only some information about the specimen that can be stably retrieved, in our case the singularities of the scattering potential, i.e. the boundaries of the molecules in the specimen.

Λ-tomography is based on the same assumptions as FBP.

Reference(s)

Local Tomography in Electron Microscopy
Quinto E T and Öktem O
Analytic methods: Λ-tomography

Basic assumptions and principle

In ET one has limited angle local tomography data. Therefore, one cannot exactly reconstruct the scattering potential of the specimen even in cases when one assumes to have a continuum of exact data!

Solution proposed by Λ-tomography

Reconstruct only some information about the specimen that can be stably retrieved, in our case the singularities of the scattering potential, i.e. the boundaries of the molecules in the specimen.

Λ-tomography is based on the same assumptions as FBP.

Reference(s)

Local Tomography in Electron Microscopy
Quinto E T and Öktem O
Analytic methods: Λ-tomography

Reconstruction scheme

Reconstruction operator: For a function $g$ defined on a parallel beam line complex specified by the curve $S \subset S^2$, define

$$\mathcal{L}_\Lambda(g) := \mathcal{P}_S^* \mathcal{D}_S^2 g$$

where $\mathcal{D}_S^2$ is a second order differentiation along the tangential direction to the curve $S$, i.e.

$$\mathcal{D}_S^2 g(\omega, x) := \left. \frac{d^2}{ds^2} g(\omega, x + s\sigma) \right|_{s=0}$$

with $\sigma$ denoting the unit tangent to $S$ at $\omega \in S$.

Rationale: If $g = \mathcal{P}(f)$, then using microlocal analysis one can show that the visible singularities of $f$ coincide with those of $\mathcal{L}_\Lambda(g)$. Moreover, the visible singularities can be precisely described and the recovery is only mildly ill-posed even in the case of missing data.
Analytic methods: Λ-tomography

Reconstruction scheme

Reconstruction operator: For a function $g$ defined on a parallel beam line complex specified by the curve $S \subset S^2$, define

$$\mathcal{L}_\Lambda(g) := \mathcal{P}_S^* \mathcal{D}_S^2 g$$

where $\mathcal{D}_S^2$ is a second order differentiation along the tangential direction to the curve $S$, i.e.

$$\mathcal{D}_S^2 g(\omega, x) := \frac{d^2}{ds^2} g(\omega, x + s\sigma)\bigg|_{s=0}$$

with $\sigma$ denoting the unit tangent to $S$ at $\omega \in S$.

Rationale: If $g = \mathcal{P}(f)$, then using microlocal analysis one can show that the visible singularities of $f$ coincide with those of $\mathcal{L}_\Lambda(g)$. Moreover, the visible singularities can be precisely described and the recovery is only mildly ill-posed even in the case of missing data.
Analytic methods: \( \Lambda \)-tomography

Visible singularities in single axis tilting

Consider the characteristic function of a ball in \( \mathbb{R}^3 \) so the set of singularities is the sphere \( S^2 \).
Analytic methods: Λ-tomography

Visible singularities in single axis tilting

Consider the characteristic function of a ball in $\mathbb{R}^3$ so the set of singularities is the sphere $S^2$. 

Beam direction (z-axis)

Tilt axis (x-axis)
Analytic methods: Λ-tomography

Visible singularities in single axis tilting

Consider the characteristic function of a ball in $\mathbb{R}^3$ so the set of singularities is the sphere $S^2$. 
Analytic methods: FBP vs. Λ-tomography

The PSF’s of Λ-tomography and FBP in single axis tilting

FBP point spread function

Λ-tomography point spread function
Analytic methods: FBP vs. Λ-tomography
Reconstruction of *in-vitro* IgG antibodies from real data

**Specimen:** *In-vitro* specimen with monoclonal IgG antibodies (molecular weight about 150 kDa).

**Data:** 200 keV TEM single axis tilt data at 1 µm under-focus with a uniform sampling of the tilt angle in $[-60^\circ, 60^\circ]$ at 2° step. The total dose is $1520 \text{ e}^-/\text{nm}^2$.

**Reconstruction:** Region of interest is a $50 \times 50 \times 50$ voxel large with a voxel size of 0.5241 nm.
Analytic methods: FBP vs. \( \Lambda \)-tomography

Reconstruction of *in-vitro* IgG antibodies from real data

**Left:** FBP with filter bandwidth \( 10 \text{ nm}^{-1} \) post-filtered to \( 10 \text{ nm} \). The region of interest has been *cut out* from the entire reconstruction volume.

**Right:** \( \Lambda \)-reconstruction based *only* on projections that passes through the region of interest. No cut out.
Analytic methods: FBP vs. Λ-tomography

Reconstruction of *in-situ* tissue sample

**Specimen:** *In-situ* specimen from a kidney tissue sample.

**Data:** 200 kV TEM single axis tilt data at 1 µm under-focus with a uniform sampling of the tilt angle in $[-60^\circ, 60^\circ]$ at $2^\circ$ step. The total dose is $1520 \text{ e}^-/\text{nm}^2$.

**Reconstruction:** Region of interest is a $200 \times 200 \times 140$ voxel large with a voxel size of 0.5241 nm.
Analytic methods: FBP vs. Λ-tomography

Reconstruction of *in-situ* tissue sample

**Left:** FBP with filter bandwidth $10 \text{ nm}^{-1}$ post-filtered to 3 nm.

**Right:** Λ-reconstruction post-filtered to 3 nm.
Variational regularisation methods

General setup

A variational regularisation method is defined as the solution to

\[
\min_{f \in X} \lambda S(f) + \Delta(T(f), g)
\]

where \( S : X \rightarrow \mathbb{R}_+ \) is the regularisation functional which enforces uniqueness and also acts as a stabiliser by incorporating a priori knowledge about the unknown, \( \Delta : H \times H \rightarrow \mathbb{R}_+ \) is the data discrepancy functional which is an appropriate measure in data space of the dissimilarity between the measured data and the data generated by the reconstruction, and \( \lambda > 0 \) is the regularisation parameter that quantifies the compromise between accuracy (resolution) and stability.
Variational regularisation methods

General setup

A variational regularisation method is defined as the solution to

$$\min_{f \in \mathcal{X}} \lambda S(f) + \triangle \left( \mathcal{T}(f), g \right)$$

where

- $S : \mathcal{X} \to \mathbb{R}^+$ is the regularisation functional which enforces uniqueness and also acts as a stabiliser by incorporating a priori knowledge about the unknown,
- $\triangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}^+$ is the data discrepancy functional which is an appropriate measure in data space of the dissimilarity between the measured data and the data generated by the reconstruction,
- and $\lambda > 0$ is the regularisation parameter that quantifies the compromise between accuracy (resolution) and stability.
Variational regularisation methods

General setup

A variational regularisation method is defined as the solution to

$$\min_{f \in \mathcal{X}} \lambda S(f) + \Delta \left( T(f), g \right)$$

where

- $S : \mathcal{X} \to \mathbb{R}^+$ is the regularisation functional which enforces uniqueness and also acts as a stabiliser by incorporating a priori knowledge about the unknown,
- $\Delta : \mathcal{H} \times \mathcal{H} \to \mathbb{R}^+$ is the data discrepancy functional which is an appropriate measure in data space of the dissimilarity between the measured data and the data generated by the reconstruction.
- and $\lambda > 0$ is the regularisation parameter that quantifies the compromise between accuracy (resolution) and stability.
Variational regularisation methods

General setup

A variational regularisation method is defined as the solution to

$$\min_{f \in \mathcal{X}} \lambda S(f) + \triangle(\mathcal{T}(f), g)$$

where

- $S: \mathcal{X} \rightarrow \mathbb{R}^+$ is the regularisation functional which enforces uniqueness and also acts as a stabiliser by incorporating \textit{a priori} knowledge about the unknown,
- $\triangle: \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}^+$ is the data discrepancy functional which is an appropriate measure in data space of the dissimilarity between the measured data and the data generated by the reconstruction.
- $\lambda > 0$ is the regularisation parameter that quantifies the compromise between accuracy (resolution) and stability.
### Variational regularisation methods

#### Common approaches

**Choice of regularisation functional:** The choice should be based on *a priori* knowledge about the unknowns.

<table>
<thead>
<tr>
<th>Prior information</th>
<th>Regularisation functional</th>
</tr>
</thead>
<tbody>
<tr>
<td>True solution is sparse in the voxel representation <em>(i.e. true solution behaves like few outstanding features in a low-amplitude background).</em></td>
<td>$S(f) = | f |<em>{L^p} = (\int</em>{\mathcal{X}}</td>
</tr>
<tr>
<td></td>
<td>$S(f) = \int_{\mathcal{X}} \left( f(x) \ln \frac{f(x)}{\rho(x)} - f(x) + \rho(x) \right) , dx$ with $\rho \in \mathcal{X}$</td>
</tr>
<tr>
<td>Gradient of the true solution is sparse in the voxel representation <em>(i.e. the true solution behaves like a step function).</em> Case with $p = 1$ is the TV-regularisation.</td>
<td>$S(f) = | \nabla f |<em>{L^p} = (\int</em>{\mathcal{X}}</td>
</tr>
<tr>
<td>True solution has edges but is smoothly varying away from edges.</td>
<td>$S(f) = \int_{\mathcal{X}}</td>
</tr>
<tr>
<td>Edges in the true solution along specific directions are better characterised than the remaining edges.</td>
<td>$S(f) = \int_{\mathcal{X}} \phi(</td>
</tr>
<tr>
<td>True solution is smooth.</td>
<td>$S(f) = | \nabla f |<em>{L^2} = \sqrt{\int</em>{\mathcal{X}}</td>
</tr>
<tr>
<td>True solution is sparse in some representation ${\phi_i}_i$, i.e. $f = \sum_i \alpha_i \phi_i$ where most $\alpha_i \approx 0$.</td>
<td>$S(f) = (\sum_i</td>
</tr>
</tbody>
</table>
Variational regularisation methods

Common approaches

**Choice of data discrepancy functional:** The choice should be based on knowledge of the probability model for the data stochasticity. Here $\| \cdot \|_{\ell^p}$ denotes the Euclidian $p$-norm in $H \simeq \mathbb{R}^m$.

<table>
<thead>
<tr>
<th>Model for data stochasticity</th>
<th>Data discrepancy functional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data is contaminated with additive Gaussian noise with zero mean and covariance matrix $\Sigma$.</td>
<td>$\Delta (T(f), g) = | T(f) - g |_{\ell^2}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta (T(f), g) = \sqrt{(T(f) - g)^t \cdot \Sigma^{-1} \cdot (T(f) - g)}$</td>
</tr>
<tr>
<td>Data is Poisson distributed with mean given by actual measurement, i.e. noise is signal dependent.</td>
<td>$\Delta (T(f), g) = \sum_{i=1}^{m} g_i \log T(f)_i - T(f)_i$</td>
</tr>
<tr>
<td>The noise in data is impulse noise in which a random portion of the data points are corrupted, e.g. salt and pepper noise.</td>
<td>$\Delta (T(f), g) = | T(f) - g |_{\ell^0}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta (T(f), g) = | T(f) - g |_{\ell^1}$</td>
</tr>
</tbody>
</table>
**Variational regularisation methods**

**Common approaches**

**Choosing the regularisation parameter(s):** Ideally, the choice is based on the model for the data stochasticity and and the particular choice of data discrepancy functional. Here, $f_{\lambda}$ denotes the regularised solution using the regularisation parameter $\lambda$ and $f_{\text{true}}$ the true (unknown) solution.

<table>
<thead>
<tr>
<th>Prior knowledge</th>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>An <em>a priori</em> estimate of the data error and/or the value of the regularisation functional is available, i.e. we now <em>a priori</em> an estimate $\epsilon &gt; 0$ such that $\Delta(T(f_{\text{true}}), g) &lt; \epsilon$ and/or an estimate $\delta &gt; 0$ such that $S(f_{\text{true}}) \leq \delta$.</td>
<td>Morozov principle: Choose $\lambda$ such that $\Delta(T(f_{\lambda}), g) \leq \epsilon$.</td>
</tr>
<tr>
<td>No <em>a priori</em> knowledge about the data error and/or the value of the regularisation functional. Let the data $g$ choose the value of regularisation parameter.</td>
<td>Miller method: Choose $\lambda$ such that $\Delta(T(f_{\lambda}), g) &lt; \epsilon$ and $S(f_{\lambda}) \leq \delta$.</td>
</tr>
</tbody>
</table>

Generalised cross-validation: Let $f_{k, \lambda} \in \mathcal{X}$ denote the regularised solution when we have removed the $k$:th component $g_k$ of the data $g$. Choose $\lambda$ in order to predict missing data values, i.e. $T(f_{k, \lambda}) \approx g_k$ by minimising $\sum_{i=1}^{m} |T(f_{k, \lambda}) - g_k|$. L-curve: Select the $\lambda$ where the log-log plot of the curve $\lambda \mapsto \left(\Delta(T(f_{\lambda}), g), S(f_{\lambda})\right)$ has highest curvature (i.e. a corner).
Classical regularisation theory

Multi-component reconstruction problems

Two-component reconstruction problem: Assume that \( g \in \mathcal{H} \) is related to \((f, c) \in \mathcal{X} \times \mathcal{Y}\) by

\[
T(f, c) = g.
\]

Recall \((f, c) \in \mathcal{X} \times \mathcal{Y}\) from \(g \in \mathcal{H}\).

In the above, \(f \in \mathcal{X}\) would typically be the unknown of primary interest and \(c \in \mathcal{Y}\) represents the unknown parameters.

Principle of component-wise reconstruction:

- \(f(c) \in \mathcal{X}\) denotes the (regularised) solution for given \(c \in \mathcal{Y}\).
- \(c(f) \in \mathcal{Y}\) is defined analogously.
- Define the intertwined sequence \(\{(f_j, c_j)\}_{j} \subset \mathcal{X} \times \mathcal{Y}\) recursively as

\[
\begin{cases} 
  c_j := c(f_{j-1}) \\
  f_j := f(c_j)
\end{cases}
\]
Classical regularisation theory

Multi-component reconstruction problems

Two-component reconstruction problem: Assume that \( g \in \mathcal{H} \) is related to \((f, c) \in \mathcal{X} \times \mathcal{Y}\) by

\[
T(f, c) = g.
\]

Recover \((f, c) \in \mathcal{X} \times \mathcal{Y}\) from \(g \in \mathcal{H}\).

In the above, \(f \in \mathcal{X}\) would typically be the unknown of primary interest and \(c \in \mathcal{Y}\) represents the unknown parameters.

Principle of component-wise reconstruction:

- \(f(c) \in \mathcal{X}\) denotes the (regularised) solution for given \(c \in \mathcal{Y}\).
- \(c(f) \in \mathcal{Y}\) is defined analogously.
- Define the intertwined sequence \(\{(f_j, c_j)\}_j \subset \mathcal{X} \times \mathcal{Y}\) recursively as

\[
\begin{cases}
  c_j := c(f_{j-1}) \\ 
  f_j := f(c_j)
\end{cases}.
\]
Each iterate in COMET is obtained by solving the variational regularisation

\[
\min_{f \in \mathcal{X}} S_\rho(f) \quad \text{subject to} \quad \left\| \mathcal{I}(f) - g \right\|_{\mathcal{H}} \leq \varepsilon
\]

where \(-S_\rho\) is given as the relative entropy w.r.t. a prior \(\rho\), i.e.

\[
S_\rho(f) := \int \left( f(x) \ln \left( \frac{f(x)}{\rho(x)} \right) - f(x) + \rho(x) \right) \, dx.
\]

\(\varepsilon\) (and the prior \(\rho\)) and can be updated in each iterate.

COMET can also handle two-component reconstruction problems, such as the one arising in ET, by component-wise reconstruction.
COMET

Overview

1. Each iterate in COMET is obtained by solving the variational regularisation

\[
\begin{aligned}
\min_{f \in \mathcal{X}} & \; S_\rho(f) \\
\| T(f) - g \|_{\mathcal{H}} & \leq \varepsilon
\end{aligned}
\]

(2)

where \(-S_\rho\) is given as the relative entropy w.r.t. a prior \(\rho\), i.e.

\[
S_\rho(f) := \int \left( f(x) \ln \left( \frac{f(x)}{\rho(x)} \right) - f(x) + \rho(x) \right) \, dx.
\]

2. \(\varepsilon\) (and the prior \(\rho\)) and can be updated in each iterate.

3. COMET can also handle two-component reconstruction problems, such as the one arising in ET, by component-wise reconstruction.
Each iterate in COMET is obtained by solving the variational regularisation

$$\begin{cases} \min_{f \in \mathcal{X}} S_\rho(f) \\ \|T(f) - g\|_{\mathcal{H}} \leq \varepsilon \end{cases}$$

where $-S_\rho$ is given as the relative entropy w.r.t. a prior $\rho$, i.e.

$$S_\rho(f) := \int \left( f(x) \ln \left( \frac{f(x)}{\rho(x)} \right) - f(x) + \rho(x) \right) dx.$$
Why do we want to update $\varepsilon$ (and the prior $\rho$)?

As mentioned before, in ET it is difficult to \textit{a priori} determine a reliable value of $\varepsilon$ (and $\rho$). The updating of $\varepsilon$ (and the prior $\rho$) enables one to adapt the initial unreliable choice of $\varepsilon$ (and prior $\rho$).

COMET has two regularisation parameters that must be set by the user.

1. $0 \leq \delta \leq 1$ regulates the updating of $\varepsilon$: $\delta = 0$ means fitting the data as close as possible and $\delta = 1$ means the data is not trusted at all.

2. $b > 0$ regulates the degree of smoothing in the updating of the prior $\rho$ (if it is to be updated): Increase $b$ to smooth the prior more.
COMET
The regularisation parameters

Why do we want to update $\varepsilon$ (and the prior $\rho$)?

As mentioned before, in ET it is difficult to \textit{a priori} determine a reliable value of $\varepsilon$ (and $\rho$). The updating of $\varepsilon$ (and the prior $\rho$) enables one to adapt the initial unreliable choice of $\varepsilon$ (and prior $\rho$).

COMET has two regularisation parameters that must be set by the user.

1. $0 \leq \delta \leq 1$ regulates the updating of $\varepsilon$: $\delta = 0$ means fitting the data as close as possible and $\delta = 1$ means the data is not trusted at all.

2. $b > 0$ regulates the degree of smoothing in the updating of the prior $\rho$ (if it is to be updated): Increase $b$ to smooth the prior more.
Why do we want to update $\varepsilon$ (and the prior $\rho$)?

As mentioned before, in ET it is difficult to \textit{a priori} determine a reliable value of $\varepsilon$ (and $\rho$). The updating of $\varepsilon$ (and the prior $\rho$) enables one to adapt the initial unreliable choice of $\varepsilon$ (and prior $\rho$).

COMET has two regularisation parameters that must be set by the user.

1. $0 \leq \delta \leq 1$ regulates the updating of $\varepsilon$: $\delta = 0$ means fitting the data as close as possible and $\delta = 1$ means the data is not trusted at all.

2. $b > 0$ regulates the degree of smoothing in the updating of the prior $\rho$ (if it is to be updated): Increase $b$ to smooth the prior more.
COMET

The iterative scheme

COMET iterates \((f_j, c_j) \in \mathcal{X} \times \mathcal{Y}\) are defined as follows:

\[
c_j := \arg\min_{c \in \mathcal{Y}} \left\| T(f_{j-1}, c) - g \right\|_{\mathcal{H}}
\]

Updating the additional unknown components by least-squares fit
COMET

The iterative scheme

COMET iterates \((f_j, c_j) \in \mathcal{X} \times \mathcal{Y}\) are defined as follows:

\[
c_j := \arg\min_{c \in \mathcal{Y}} \left\| \mathcal{T}(f_{j-1}, c) - g \right\|_H
\]

\[
\rho_j := \begin{cases} 
\mathcal{F}_b(f_{j-1}) & \text{if prior is to be updated,} \\
\rho & \text{if prior is not updated,}
\end{cases}
\]

Updating the prior, \(\mathcal{F}_b(\cdot)\) is a smoothing operator e.g. a low-pass filtering with \(b > 0\) denoting the cut-off in the frequency domain.
COMET

The iterative scheme

COMET iterates \((f_j, c_j) \in \mathcal{X} \times \mathcal{Y}\) are defined as follows:

\[
c_j := \arg\min_{c \in \mathcal{Y}} \left\| \mathcal{T}(f_{j-1}, c) - g \right\|_{\mathcal{H}}
\]

\[
\rho_j := \begin{cases} 
\mathcal{F}_b(f_{j-1}) & \text{if prior is to be updated,} \\
\rho & \text{if prior is not updated,}
\end{cases}
\]

\[
\varepsilon_j := \varepsilon_{\min}(g, c_j) + \delta \left( \left\| \mathcal{T}(\rho_j, c_j) - g \right\|_{\mathcal{H}} - \varepsilon_{\min}(g, c_j) \right)
\]

Updating the estimate of the data error,

\[
\varepsilon_{\min}(g, c) := \inf_{f \in \mathcal{X}} \left\| \mathcal{T}(f, c) - g \right\|_{\mathcal{H}}
\]
COMET

The iterative scheme

COMET iterates \((f_j, c_j) \in \mathcal{X} \times \mathcal{Y}\) are defined as follows:

\[
c_j := \arg\min_{c \in \mathcal{Y}} \| T(f_{j-1}, c) - g \|_H
\]

\[
\rho_j := \begin{cases} 
F_b(f_{j-1}) & \text{if prior is to be updated,} \\
\rho & \text{if prior is not updated,}
\end{cases}
\]

\[
\varepsilon_j := \varepsilon_{\min}(g, c_j) + \delta \left( \| T(\rho_j, c_j) - g \|_H - \varepsilon_{\min}(g, c_j) \right)
\]

\[
f_j := \begin{cases} 
\arg\min_{f \in \mathcal{X}} S_{\rho_j}(f) & \text{if prior is to be updated,} \\
\arg\min_{f \in \mathcal{X}} \| T(f, c_j) - g \|_H \leq \varepsilon_j.
\end{cases}
\]

Updating the object that is of primary interest by entropy regularisation
Setting the regularisation parameters

By empirical experience! For data with good signal-to-noise ratio (typically \textit{in-situ} data), \( \delta \approx 0.5 \) and \( b \approx 6 \text{ nm}^{-1} \). For data with poor signal-to-noise ratio (typically \textit{in-vitro} data), \( \delta \approx 0.8 \) and \( b \approx 10 \text{ nm}^{-1} \).

- Ideally, the regularisation parameters \( \delta \) and \( b \) should be set using knowledge about the stochasticity of the data and an estimate of the data error.
Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: a test application with Adenovirus
Skoglund U, Öfverstedt, L-G, Burnett R M and Bricogne G.

First publication describing usage of entropy regularisation in ET. Does not include a description of the updating of $\varepsilon$ and $\rho$, since at that time those parameters were constant and assumed to be known a priori.

A component-wise iterated relative entropy regularisation method with updated prior and regularisation parameter
Rullgård H, Öktem O and Skoglund U.

The formal mathematical analysis and description of COMET.
In $\ell_1$-regularisation the reconstruction problem is defined by solving

$$\min_{f \in \mathcal{X}} S_\lambda(f) + \| T(f) - g \|_{\mathcal{H}}$$

where the regularisation functional $S_\lambda$ can be defined as

$$S_\lambda(f) := \lambda_1 \| f \|_{\mathcal{L}^1} + \lambda_2 \| \nabla f \|_{\mathcal{L}^1}$$

for $\lambda := (\lambda_1, \lambda_2)$ which is the regularisation parameter for the method. In the above expression, $\| \cdot \|_{\mathcal{L}^1}$ denotes the $\mathcal{L}^1$-norm, i.e.

$$\| f \|_{\mathcal{L}^1} := \int |f(x)| \, dx \quad \text{and} \quad \| \nabla f \|_{\mathcal{L}^1} := \int |\nabla f(x)| \, dx$$

and second term to the right is often referred to as the total variation of $f$. 

\[ \ell_1\text{-TV regularisation} \]
\textbf{$\ell_1$-TV regularisation}

In $\ell_1$-regularisation the reconstruction problem is defined by solving

$$
\min_{f \in \mathcal{X}} S_\lambda(f) + \left\| T(f) - g \right\|_H
$$

where the regularisation functional $S_\lambda$ can be defined as

$$
S_\lambda(f) := \lambda_1 \left\| f \right\|_{L^1} + \lambda_2 \left\| \nabla f \right\|_{L^1}
$$

for $\lambda := (\lambda_1, \lambda_2)$ which is the regularisation parameter for the method. In the above expression, $\left\| \cdot \right\|_{L^1}$ denotes the $L^1$-norm, \textit{i.e.}

$$
\left\| f \right\|_{L^1} := \int |f(x)| \, dx \quad \text{and} \quad \left\| \nabla f \right\|_{L^1} := \int |\nabla f(x)| \, dx
$$

and second term to the right is often referred to as the total variation of $f$. 
For efficient removal of background noise, the regularisation parameter(s) $\lambda$ should be chosen so that for a small selection of “test functions” $h_{\text{test}}$, e.g. characteristic functions of balls,

$$2 < \frac{S_\lambda(h_{\text{test}})}{\sigma(h_{\text{test}})} < 5$$

where $\sigma(h_{\text{test}}) := \text{Variance} \left[ \langle T(h_{\text{test}}), E \rangle_H \right]^{1/2}$ holds. In the above, $E$ is the random variable representing the noise component of the data.

- In ET, $\sigma(h_{\text{test}})$ can be reliably estimated directly from the tilt series.
- Finally, for a given choice of $\lambda$, one can also estimate the size of features that can be recovered.
\textbf{\(\ell_1\)-TV regularisation}

\textbf{Specific properties}

- For efficient removal of background noise, the regularisation parameter(s) \(\lambda\) should be chosen so that for a small selection of “test functions” \(h_{\text{test}}\), e.g. characteristic functions of balls,

\[
2 < \frac{S_\lambda(h_{\text{test}})}{\sigma(h_{\text{test}})} < 5 \quad \text{where} \quad \sigma(h_{\text{test}}) := \text{Variance} \left[ \left\langle T(h_{\text{test}}), E \right\rangle_H \right]^{1/2}
\]

holds. In the above, \(E\) is the random variable representing the noise component of the data.

- In ET, \(\sigma(h_{\text{test}})\) can be reliably estimated directly from the tilt series.

- Finally, for a given choice of \(\lambda\), one can also estimate the size of features that can be recovered.
\( \ell_1 \text{-TV regularisation} \)

Specific properties

- For efficient removal of background noise, the regularisation parameter(s) \( \lambda \) should be chosen so that for a small selection of “test functions” \( h_{\text{test}} \), e.g. characteristic functions of balls,

\[
2 < \frac{S_\lambda(h_{\text{test}})}{\sigma(h_{\text{test}})} < 5 \quad \text{where} \quad \sigma(h_{\text{test}}) := \text{Variance} \left[ \left< \mathcal{T}(h_{\text{test}}), E \right> \mathcal{H} \right]^{1/2}
\]

holds. In the above, \( E \) is the random variable representing the noise component of the data.

- In ET, \( \sigma(h_{\text{test}}) \) can be reliably estimated directly from the tilt series.

- Finally, for a given choice of \( \lambda \), one can also estimate the size of features that can be recovered.
A new principle for choosing regularization parameter in certain inverse problems
Rullgård H

A parameter choice rule (i.e. method for choosing the regularisation parameter) is developed for total variation regularisation of reconstruction problems with high levels of noise in the data. The parameter choice rule is then applied to the reconstruction problem in electron tomography.
Tests on data
Comparison of various methods
Reconstruction of *in-vitro* RNA polymerase II molecules from simulated data

**Specimen:** *In-vitro* specimen with one RNA polymerase II molecule (with a “diameter” of about 13 nm) in a 50 nm thick slab.

**Data:** Simulated 200 kV TEM single axis tilt data at 24900× magnification with 1 µm under-focus. Uniform sampling of the tilt angle in $[-60^\circ, 60^\circ]$ at 1° step. The total dose is $2000 \, \text{e}^-/\text{nm}^2$.

**Reconstruction:** Region of interest is a $200 \times 200 \times 200$ voxel large with a voxel size of 0.5622 nm.
Comparison of various methods
Slice through reconstruction

Phantom

FBP

SIRT

COMET

$\ell_1$-TV regularisation

Size of scale-bar is 10 nm.
Statistical reconstruction methods

Main principle

The philosophy is to recast the inverse problem in the form of statistical quest for information:

Classical reconstruction methods: Reconstructs a single estimate of the unknowns and seeks to answer the question:
What is the value of the unknowns?

Statistical reconstruction methods: Reconstructs the probability distribution of the unknowns and seeks to answer the question:
What is our information about the unknowns?
Statistical reconstruction methods

Main principle

The philosophy is to recast the inverse problem in the form of statistical quest for information:

**Classical reconstruction methods:** Reconstructs a single estimate of the unknowns and seeks to answer the question:

*What is the value of the unknowns?*

**Statistical reconstruction methods:** Reconstructs the probability distribution of the unknowns and seeks to answer the question:

*What is our information about the unknowns?*
Statistical reconstruction methods

Common approaches

- Maximum Likelihood
- Minimum Inaccuracy
- Probability Distribution Matching
- Minimising measures of information content, e.g. Maximum Entropy methods
- Bayesian inference
Bayesian inference
Reformulation of the reconstruction problem

The abstract reconstruction problem

We are given random variables $G, F,$ and $E$ with values in $\mathcal{H}, \mathcal{X},$ and $\mathcal{Y},$ respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}},$
- and the prior distribution $\pi_{\text{prior}}$.

$G$ is the random variable with values in data space $\mathcal{H}$ modelling the measured data $g \in \mathcal{H}$.
Bayesian inference
Reformulation of the reconstruction problem

The abstract reconstruction problem

We are given random variables $G$, $F$, and $E$ with values in $\mathcal{H}$, $\mathcal{X}$, and $\mathcal{Y}$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}}$,
- and the prior distribution $\pi_{\text{prior}}$.

$F$ is the random variable with values in reconstruction space $\mathcal{X}$ modelling all possible solutions.
Bayesian inference
Reformulation of the reconstruction problem

The abstract reconstruction problem
We are given random variables $G$, $F$, and $E$ with values in $\mathcal{H}$, $\mathcal{X}$, and $\mathcal{Y}$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from
- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}}$,
- and the prior distribution $\pi_{\text{prior}}$.

$E$ is the random variable with values in $\mathcal{Y}$ modelling additional noise components in the data that are independent of $F$, typically additive or multiplicative noise. Often, $\mathcal{Y} = \mathcal{H}$. 
Bayesian inference
Reformulation of the reconstruction problem

The abstract reconstruction problem

We are given random variables $G, F, \text{ and } E$ with values in $\mathcal{H}, \mathcal{X}, \text{ and } \mathcal{Y}$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}}$,
- and the prior distribution $\pi_{\text{prior}}$.

Model: Typical examples are $G = \mathcal{T}(F) + E$ or $G = C + E$ where the probability distribution of $C$ depends on $\mathcal{T}(F)$. Here, $\mathcal{T} : \mathcal{X} \rightarrow \mathcal{H}$ denotes the forward operator.
Bayesian inference

Reformulation of the reconstruction problem

The abstract reconstruction problem

We are given random variables $G$, $F$, and $E$ with values in $\mathcal{H}$, $\mathcal{X}$, and $\mathcal{Y}$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{post}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{likelihood}$,
- and the prior distribution $\pi_{prior}$.

**Posterior distribution:** $\pi_{post}(f) := \pi(f|G = g)$ is the conditional probability distribution of $F$ given $G$ and describes all possible solutions and their probabilities.
Bayesian inference
Reformulation of the reconstruction problem

The abstract reconstruction problem
We are given random variables $G, F, \text{ and } E$ with values in $\mathcal{H}, X, \text{ and } Y$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}}$, and
- and the prior distribution $\pi_{\text{prior}}$.

Likelihood function: $\pi_{\text{likelihood}}(g) := \pi(g|F = f)$ is the conditional probability distribution of $G$ given $F$ that models the stochasticity of the data.
Bayesian inference

Reformulation of the reconstruction problem

The abstract reconstruction problem

We are given random variables $G$, $F$, and $E$ with values in $\mathcal{H}$, $\mathcal{X}$, and $\mathcal{Y}$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}}$,
- and the prior distribution $\pi_{\text{prior}}$.

Prior distribution: $f \mapsto \pi_{\text{prior}}(f)$ is the probability distribution of $F$ representing the a priori knowledge we have about the solution to the reconstruction problem.
Bayesian inference
Reformulation of the reconstruction problem

The abstract reconstruction problem

We are given random variables $G$, $F$, and $E$ with values in $\mathcal{H}$, $\mathcal{X}$, and $\mathcal{Y}$, respectively. Next, we have a model relating the probability distribution of $G$ to the probability distributions of $\mathcal{T}(F)$ and $E$. Recover the posterior distribution $\pi_{\text{post}}$ from

- a sample $g \in \mathcal{H}$ of $G$,
- the likelihood function $\pi_{\text{likelihood}}$,
- and the prior distribution $\pi_{\text{prior}}$.

From now on we assume that the reconstruction space is discretised, i.e. $\mathcal{X} \simeq \mathbb{R}^n$ for some $n$ and that $\mathcal{Y} = \mathcal{H}$ (keep in mind that $\mathcal{H} \simeq \mathbb{R}^m$ where $m$ is the number of data points).
Bayesian inference
Inverse Problems and Bayes’ Formula

Bayes’ Formula

Assume that for the given measured data $g \in \mathcal{H}$, the marginal distribution of $G$ at $g$ is positive, i.e. $\pi_G(g) > 0$. Then

$$
\pi_{\text{post}}(f) = \frac{\pi_{\text{prior}}(f) \pi_{\text{likelihood}}(g)}{\pi_G(g)}.
$$

Note: The marginal distribution $\pi_G$ plays the role of a norming constant and is usually of little importance. If $\pi_G(g) = 0$, then we have measurement data $g$ that have zero probability, so the underlying models are not consistent with the reality.
Bayesian inference
Inverse Problems and Bayes’ Formula

Bayes’ Formula
Assume that for the given measured data $g \in \mathcal{H}$, the marginal distribution of $G$ at $g$ is positive, i.e. $\pi_G(g) > 0$. Then

$$\pi_{\text{post}}(f) = \frac{\pi_{\text{prior}}(f) \pi_{\text{likelihood}}(g)}{\pi_G(g)}.$$

Solving an inverse problem may therefore be broken into three subtasks:

1. Construction of likelihood function
2. Construction of prior
3. Construction of estimators that allow us to explore the posterior probability distribution $f \mapsto \pi_{\text{post}}(f)$. 
## Bayesian inference

### Construction of the likelihood function

<table>
<thead>
<tr>
<th>Model</th>
<th>Likelihood function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent additive noise: $G = \mathcal{I}(F) + E$ with $F$ and $E$ independent and the probability distribution $\pi_{\text{noise}}$ of the noise $E$ is known.</td>
<td>$\pi_{\text{likelihood}}(g) = \pi_{\text{noise}}(g - \mathcal{I}(f))$</td>
</tr>
<tr>
<td>Dependent additive noise: $G = \mathcal{I}(F) + E$ with $F$ and $E$ dependent and the conditional probability distribution $\pi_E$ of the noise $E$ given $F = f$ is known.</td>
<td>$\pi_{\text{likelihood}}(g) = \pi_E(g - \mathcal{I}(f))$</td>
</tr>
<tr>
<td>Multiplicative noise: $G = \mathcal{I}(F)E$ (component-wise multiplication) where $E$ is mutually independent of $F$ and its components $E_i$ are independent with known probability distribution $\pi_{\text{noise}}^i$.</td>
<td>$\pi_{\text{likelihood}}(g) = \prod_{i=1}^{m} \frac{1}{\mathcal{I}(f)<em>i} \pi</em>{\text{noise}}^i \left( \frac{g_i}{\mathcal{I}(f)_i} \right)$</td>
</tr>
<tr>
<td>Counting process data: $G \sim \text{Poisson}(\mathcal{I}(f))$ for $f \in \mathcal{X}$.</td>
<td>$\pi_{\text{likelihood}}(g) = \prod_{i=1}^{m} \frac{\mathcal{I}(f)_i^{g_i}}{g_i!} \exp \left( -\mathcal{I}(f)_i \right)$</td>
</tr>
<tr>
<td>Counting process data with additive noise: The counting observation $C$ (which is itself a random variable) is corrupted by independent additive noise, i.e. $G = C + E$ where $C \sim \text{Poisson}(\mathcal{I}(f))$ for $f \in \mathcal{X}$ and the components $E_i$ of $E$ are independent with known probability distribution $\pi_{\text{noise}}^i$.</td>
<td>$\pi_{\text{likelihood}}(g) = \prod_{i=1}^{m} \sum_{k=0}^{\infty} \pi_{\text{noise}}^i \left( \frac{\mathcal{I}(f)_i^{k} (g_i - k) \mathcal{I}(f)_i^{k}}{k!} \right) \exp \left( -\mathcal{I}(f)_i \right)$</td>
</tr>
</tbody>
</table>
Bayesian inference
Construction of the prior

**Challenge:** Often, prior knowledge of the unknown is *qualitative*. This must be transformed into *quantitative* form which then can be encoded into the prior density.

**Desirable property:** Should assign low probability to un-expectable solutions and high probability to expectable solutions.
Bayesian inference

Construction of the prior

Gaussian priors

Prior distribution: \( F \sim \mathcal{N}(f_0, \Sigma_{\text{prior}}) \) i.e. \( F \) has a Gaussian distribution with mean \( f_0 \in \mathcal{X} \) and covariance matrix \( \Sigma_{\text{prior}} \):

\[
\pi_{\text{prior}}(f) \propto \exp \left( -\frac{1}{2} (f - f_0)^t \cdot \Sigma_{\text{prior}}^{-1} \cdot (f - f_0) \right)
\]

Prior information: Requires a priori knowledge of the mean \( f_0 \in \mathcal{X} \) and covariance matrix \( \Sigma_{\text{prior}} \).

Advantages: Gaussian priors are relatively easy to handle (can work with closed form expressions). Next, due to the central limit theorem, Gaussian priors are applicable also for \( F \) that is non-Gaussian when the observation of \( F \) is physically based on a large number of mutually independent random events.
Bayesian inference

Construction of the prior

Gibbs priors

Prior distribution: \( \pi_{\text{prior}}(f) \propto \exp(-\mu S(f)) \).

Prior information: Requires \textit{a priori} knowledge of \( \mu > 0 \) and the functional \( S: \mathcal{H} \rightarrow \mathbb{R}^+ \).

Advantages: Appropriate choice of \( S \) allows us to create priors that encode the same \textit{a priori} information as in classical regularisation theory.

Disadvantages: \( f \mapsto \pi_{\text{prior}}(f) \) can be interpreted as a proper probability density distribution only under certain circumstances, \textit{e.g.} if \( S(f) = \|L \cdot f\|_{\ell^2} \) then \( \ker L = \{0\} \) needs to hold.
Bayesian inference

Construction of the prior

Impulse priors

Prior distribution(s): Let $\mu > 0$ be fixed:

$$
\pi_{\text{prior}}(f) \propto \exp(-\mu \| f \|_{\ell^p}) \quad \text{for } 0 < p \leq 1
$$

$$
\pi_{\text{prior}}(f) \propto \mu^n \prod_{i=1}^{n} \frac{1}{1 + \mu^2 f_i^2}
$$

$$
\pi_{\text{prior}}(f) \propto \exp\left(-\mu \sum_{i=1}^{n} \left( f_i \ln \frac{f_i}{\rho_j} - f_i + \rho_j \right)\right) \quad \text{for } \rho \in \mathcal{X}
$$

$$
\pi_{\text{prior}}(f) \propto \prod_{i=1}^{n} \frac{1}{f_j \sigma_j} \exp\left(-\frac{1}{2\sigma_j^2} (\ln f_j - \ln \rho_j)^2\right) \quad \text{for } \sigma_j > 0, \rho \in \mathcal{X}
$$

Advantages: Appropriate when the true solution has low average amplitude with few outliers standing out against the background.
Bayesian inference

Construction of the prior

Markov Random Fields

Prior distribution(s): Assume we are given functions $V_j: \mathcal{X} \rightarrow \mathbb{R}$ that depend only on $f_j$ and components $f_k$ with indices $k$ in a neighbourhood of $j$. Then we define

$$
\pi_{\text{prior}}(f) \propto \exp\left(-\mu \sum_{j=1}^{n} V_j(f)\right)
$$

where $V_j$ specifies the correlation between $f_j$ and the remaining components $f_k$ of $f$.

Advantages: Appropriate when the true solution has a fairly known correlation between various components.
Bayesian inference

Estimators

- Maximum a posteriori (MAP)
- Conditional mean
- Marginal MAP

MAP needs optimisation algorithms, Conditional mean needs integration methods in $\mathcal{X}$, and Marginal MAP needs integration and optimisation.

MAP is defined as

$$f_{\text{MAP}} := \arg\max_{f \in \mathcal{X}} \pi_{\text{post}}(f)$$

i.e. $f_{\text{MAP}}$ is the estimator that is the most probable given the data $g$.

Compare with the maximum likelihood estimate

$$f_{\text{ML}} := \arg\max_{f \in \mathcal{X}} \pi_{\text{likelihood}}(f)$$

which gives the estimator that is most likely to produce the data $g$. 
Bayesian inference

Estimators

- Maximum a posteriori (MAP)
- Conditional mean
- Marginal MAP

MAP needs optimisation algorithms, Conditional mean needs integration methods in $\mathcal{X}$, and Marginal MAP needs integration and optimisation.

MAP is defined as

$$f_{\text{MAP}} := \arg\max_{f \in \mathcal{X}} \pi_{\text{post}}(f)$$

i.e. $f_{\text{MAP}}$ is the estimator that is the most probable given the data $g$.

Compare with the maximum likelihood estimate

$$f_{\text{ML}} := \arg\max_{f \in \mathcal{X}} \pi_{\text{likelihood}}(f)$$

which gives the estimator that is most likely to produce the data $g$. 
Bayesian inference

Relation to classical regularisation theory

- Assume \( F \) has a Gibbs prior, i.e. \( \pi_{\text{prior}}(f) \propto \exp(-\mu S(f)) \).
- Introduce the negative logarithm (the so-called neglog) of the likelihood and the prior. Then the corresponding neglog of the posterior probability distribution can be written as

\[
    f \mapsto \lambda S(f) + \Delta(g, T(f))
\]

where \( \lambda > 0 \) is some constant (depending on \( \mu \)) and

\[
    \Delta(g, T(f)) := -C_1 \ln(\pi_{\text{likelihood}}(g)) + C_2
\]

for suitable constants \( C_1, C_2 \) that can be introduced in order to have the most simple expression.

- The maximisation in MAP is now transformed into a minimisation

\[
    f_{\text{MAP}} := \arg\max_{f \in \mathcal{X}} \pi_{\text{post}}(f) = \arg\min_{f \in \mathcal{X}} \lambda S(f) + \Delta(g, T(f)).
\]
Bayesian inference

Relation to classical regularisation theory

- Assume $F$ has a Gibbs prior, i.e. $\pi_{\text{prior}}(f) \propto \exp(-\mu S(f))$.
- Introduce the negative logarithm (the so-called neglog) of the likelihood and the prior. Then the corresponding neglog of the posterior probability distribution can be written as

$$f \mapsto \lambda S(f) + \Delta(g, T(f))$$

where $\lambda > 0$ is some constant (depending on $\mu$) and

$$\Delta(g, T(f)) := -C_1 \ln(\pi_{\text{likelihood}}(g)) + C_2$$

for suitable constants $C_1, C_2$ that can be introduced in order to have the most simple expression.

- The maximisation in MAP is now transformed into a minimisation

$$f_{\text{MAP}} := \underset{f \in \mathcal{X}}{\text{argmax}} \pi_{\text{post}}(f) = \underset{f \in \mathcal{X}}{\text{argmin}} \lambda S(f) + \Delta(g, T(f)).$$
Bayesian inference

Relation to classical regularisation theory

- Assume $F$ has a Gibbs prior, i.e. $\pi_{\text{prior}}(f) \propto \exp(-\mu S(f))$.
- Introduce the negative logarithm (the so-called neglog) of the likelihood and the prior. Then the corresponding neglog of the posterior probability distribution can be written as

$$f \mapsto \lambda S(f) + \triangle(g, T(f))$$

where $\lambda > 0$ is some constant (depending on $\mu$) and

$$\triangle(g, T(f)) := -C_1 \ln(\pi_{\text{likelihood}}(g)) + C_2$$

for suitable constants $C_1, C_2$ that can be introduced in order to have the most simple expression.

- The maximisation in MAP is now transformed into a minimisation

$$f_{\text{MAP}} := \arg\max_{f \in \mathcal{X}} \pi_{\text{post}}(f) = \arg\min_{f \in \mathcal{X}} \lambda S(f) + \triangle(g, T(f)).$$
Advantages of Bayesian inference vs. classical regularisation theory

- Explicit account of the errors and noise.
- A large class of priors via explicit or implicit modelling, e.g. more specific and specialised priors can be created through introduction of hidden variables.
- A coherent approach to combine information content of the data and priors.
- Selection of regularisation parameter (which in the Bayesian case are the unknown parameters in the prior) and multi-component reconstruction problems can be dealt with systematically by hyperparameters.
- Image processing operations, such as segmentation and labelling, can be included in the reconstruction by hidden variables.
Acknowledgments

Mathematics research
Dr. Hans Rullgård, Department of Mathematics, Stockholm University
Prof. Jan Boman, Department of Mathematics, Stockholm University
Prof. Jan-Olov Strömberg, Department of Mathematics, Royal Institute of Technology
Prof. Todd Quinto, Department of Mathematics, Tufts University
Prof. Dr. Alfred K. Louis, Numerische Mathematik, Universität des Saarlandes
Prof. Ali Mohammad-Djafari, CNRS, Paris
Prof. Peter Maass, Zentrum für Technomathematik, Universität Bremen

Physics research & software development
Dr. Duccio Fanelli, Dipartimento di Energetica ‘Sergio Stecco’, University of Florence
Dr. Hans Rullgård, Department of Mathematics, Stockholm University
Scientists at Sidec Research & Development

Electron tomography
Prof. Ulf Skoglund, Department of Cell and Molecular Biology, Karolinska Institutet
Dr. Sara Sandin, MRC, Cambridge
Prof. Roger Kornberg, Stanford University Medical School, Stanford
Scientists at Sidec Protein Tomography lab.