# $L^1$ regularization in electron tomography

Hans Rullgård

February 12, 2008

# Electron tomography

- Electron tomography (ET) is a tomographic method that can be used for 3D imaging of biological macromolecules, structures in cells etc.

# Electron tomography

- Electron tomography (ET) is a tomographic method that can be used for 3D imaging of biological macromolecules, structures in cells etc.
- The specimen is in the form of a thin layer ($\sim 100\ \mathrm{nm}$ thick), cooled by liquid nitrogen.

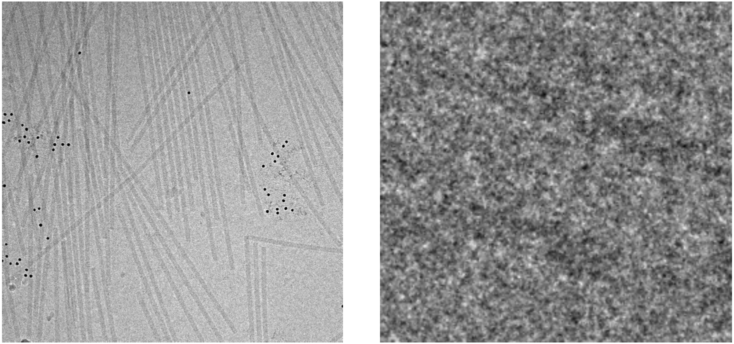
# Electron tomography

- Electron tomography (ET) is a tomographic method that can be used for 3D imaging of biological macromolecules, structures in cells etc.
- The specimen is in the form of a thin layer ($\sim 100$ nm thick), cooled by liquid nitrogen.
- A series of images, a tilt series, is collected in a transmission electron microscope. Due to the shape of the specimen, the tilt range is typically restricted to $\pm 60^\circ$. We have a limited angle problem.

# Electron tomography

- Electron tomography (ET) is a tomographic method that can be used for 3D imaging of biological macromolecules, structures in cells etc.
- The specimen is in the form of a thin layer ($\sim 100\ \mathrm{nm}$ thick), cooled by liquid nitrogen.
- A series of images, a tilt series, is collected in a transmission electron microscope. Due to the shape of the specimen, the tilt range is typically restricted to $\pm 60^\circ$. We have a limited angle problem.
- The specimen is damaged by too much radiation. This sets a limit to the total electron dose that can be used. The images collected are very noisy.

# ET images



A high dose ET image (left) and a part of an image from the tilt series (right) of a specimen containing Tobacco Mosaic Virus (TMV).

# The forward operator in ET

- A common assumption is that the electrons scatter against the electrostatic potential $f$ of the specimen. An image is obtained by projecting $f$ onto a plane perpendicular to the optical axis of the microscope.

# The forward operator in ET

- A common assumption is that the electrons scatter against the electrostatic potential $f$ of the specimen. An image is obtained by projecting $f$ onto a plane perpendicular to the optical axis of the microscope.
- The projected image is convolved with a point spread function. The forward operator $T$ is defined as projection followed by convolution.

# The forward operator in ET

- A common assumption is that the electrons scatter against the electrostatic potential $f$ of the specimen. An image is obtained by projecting $f$ onto a plane perpendicular to the optical axis of the microscope.
- The projected image is convolved with a point spread function. The forward operator $T$ is defined as projection followed by convolution.
- Each pixel of the recorded images is a random variable whose expected value is (approximately) given by $T(f)$.

# Inverse problem in general

Forward operator $T : X \to Y$.
Assume $T$ is linear, $Y$ is a Hilbert space.

# Inverse problem in general

Forward operator $T : X \to Y$.
Assume $T$ is linear, $Y$ is a Hilbert space.

- $f^{\mathrm{true}} \in X$ is unknown.
- Measure $g = T(f^{\mathrm{true}}) + g^{\mathrm{noise}} \in Y$.
- $g^{\mathrm{noise}}$ is a sample of a random vector $G^{\mathrm{noise}} \in Y$, $\mathrm{E}[G^{\mathrm{noise}}] = 0$.

# Inverse problem in general

Forward operator $T : X \to Y$.
Assume $T$ is linear, $Y$ is a Hilbert space.

- $f^{\mathrm{true}} \in X$ is unknown.
- Measure $g = T(f^{\mathrm{true}}) + g^{\mathrm{noise}} \in Y$.
- $g^{\mathrm{noise}}$ is a sample of a random vector $G^{\mathrm{noise}} \in Y$, $\mathrm{E}[G^{\mathrm{noise}}] = 0$.

Inverse problem: Find $f \in X$ approximating $f^{\mathrm{true}}$.

# Variational regularization

Let

$$f^{\mathrm{rec}} = S_\lambda(g) := \arg\min_{f \in X} R_\lambda(f) + \frac{1}{2}\|W^{1/2}(T(f) - g)\|_Y^2$$

$W$ is a self-adjoint positive definite operator on $Y$.
$R_\lambda : X \to \mathbb{R}$ is a *regularization functional* depending on parameters $\lambda$.
For suitable choice of $\lambda$, we hope that $f^{\mathrm{rec}}$ will be a good approximation of $f^{\mathrm{true}}$.

# $L^1$ methods

## Definition

Let us say that $R_\lambda$ is of $L^1$ type if it satisfies the following:

1. Convexity: $R_\lambda(f)$ is a convex function of $f$ for each $\lambda$.
2. Homogeneity: $R_\lambda(\alpha f) = |\alpha| R_\lambda(f)$ for $\alpha \in \mathbb{R}$.
3. Additivity: $R_\lambda(f_1 + f_2) = R_\lambda(f_1) + R_\lambda(f_2)$ if $f_1$ and $f_2$ have disjoint support.

# $L^1$ methods

## Definition

Let us say that $R_\lambda$ is of $L^1$ type if it satisfies the following:

1. Convexity: $R_\lambda(f)$ is a convex function of $f$ for each $\lambda$.
2. Homogeneity: $R_\lambda(\alpha f) = |\alpha| R_\lambda(f)$ for $\alpha \in \mathbb{R}$.
3. Additivity: $R_\lambda(f_1 + f_2) = R_\lambda(f_1) + R_\lambda(f_2)$ if $f_1$ and $f_2$ have disjoint support.

## Example

The total variation is an example of an $L^1$ regularization functional:

$$R_\lambda(f) = \lambda \int_\Omega |\nabla f(x)|\, dx.$$

# Goals of regularization theory

A theory of $L^1$ regularization should help us answer questions such as

- How should the regularization parameter be chosen?

# Goals of regularization theory

A theory of $L^1$ regularization should help us answer questions such as

- How should the regularization parameter be chosen?
- How much noise would be expected in the reconstruction?

# Goals of regularization theory

A theory of $L^1$ regularization should help us answer questions such as

- How should the regularization parameter be chosen?
- How much noise would be expected in the reconstruction?
- How certain can we be that a feature in reconstruction really comes from $f^{\mathrm{true}}$?

# Goals of regularization theory

A theory of $L^1$ regularization should help us answer questions such as

- How should the regularization parameter be chosen?
- How much noise would be expected in the reconstruction?
- How certain can we be that a feature in reconstruction really comes from $f^{\mathrm{true}}$?
- Which features in $f^{\mathrm{true}}$ can we hope to reconstruct?

# A simpler optimization problem

To help answer these questions, I will consider a much simpler optimization problem:

$$\alpha_\lambda(f,g) := \arg\min_{\alpha\in\mathbb{R}} R_\lambda(\alpha f) + \frac{1}{2}\|W^{1/2}(T(\alpha f) - g)\|^2.$$

# A simpler optimization problem

To help answer these questions, I will consider a much simpler optimization problem:

$$\alpha_\lambda(f,g) := \arg\min_{\alpha\in\mathbb{R}} R_\lambda(\alpha f) + \frac{1}{2}\|W^{1/2}(T(\alpha f) - g)\|^2.$$

The solution can be computed explicitly:

$$\alpha_\lambda(f,g) = \begin{cases} \frac{\langle Wg, T(f)\rangle - R_\lambda(f)}{\|W^{1/2}T(f)\|^2} & \text{if } \langle Wg, T(f)\rangle > R_\lambda(f) \\ 0 & \text{if } |\langle Wg, T(f)\rangle| \leq R_\lambda(f) \\ \frac{\langle Wg, T(f)\rangle + R_\lambda(f)}{\|W^{1/2}T(f)\|^2} & \text{if } \langle Wg, T(f)\rangle < -R_\lambda(f) \end{cases}.$$

# Main hypothesis

Knowing something about $\alpha_\lambda(f, g)$ for a selection of "test functions" $f$ can tell us something about $S_\lambda(g)$.

# Connection between $S_\lambda(g)$ and $\alpha_\lambda(f, g)$

To find rigorous connections between $S_\lambda(g)$ and $\alpha_\lambda(f, g)$ is not so easy. At least we have the following:

**Lemma**

*$S_\lambda(g) = 0$ if and only if $\alpha_\lambda(f, g) = 0$ for all $f \in X$.*

# Stochastics of the noise

I want to choose a set of test functions $f$ and look at the stochastic properties of $\alpha_\lambda(f, G^{\mathrm{noise}})$.

# Stochastics of the noise

I want to choose a set of test functions $f$ and look at the stochastic properties of $\alpha_\lambda(f, G^{\mathrm{noise}})$.
Define for any $f \in X$:

$$\sigma(f) = \mathrm{Var}[\langle W G^{\mathrm{noise}}, T(f)\rangle]^{1/2}.$$

If $R_\lambda(f) = 0$ we would have

$$\mathrm{Var}[\alpha_\lambda(f, G^{\mathrm{noise}})]^{1/2} = \frac{\sigma(f)}{\|W^{1/2} T(f)\|^2}.$$

# Stochastics of the noise

I want to choose a set of test functions $f$ and look at the stochastic properties of $\alpha_\lambda(f, G^{\text{noise}})$.
Define for any $f \in X$:

$$\sigma(f) = \text{Var}[\langle WG^{\text{noise}}, T(f)\rangle]^{1/2}.$$

If $R_\lambda(f) = 0$ we would have

$$\text{Var}[\alpha_\lambda(f, G^{\text{noise}})]^{1/2} = \frac{\sigma(f)}{\|W^{1/2}T(f)\|^2}.$$

Recall that $\alpha_\lambda(f, g) = 0$ iff $|\langle Wg, T(f)\rangle| \leq R_\lambda(f)$. For significant reduction of the variance of $\alpha_\lambda(f, g)$ we should have $s_\lambda(f) := R_\lambda(f)/\sigma(f) \gg 1$.

# Stochastics of the noise

I want to choose a set of test functions $f$ and look at the stochastic properties of $\alpha_\lambda(f, G^{\mathrm{noise}})$.
Define for any $f \in X$:

$$\sigma(f) = \mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]^{1/2}.$$

If $R_\lambda(f) = 0$ we would have

$$\mathrm{Var}[\alpha_\lambda(f, G^{\mathrm{noise}})]^{1/2} = \frac{\sigma(f)}{\|W^{1/2}T(f)\|^2}.$$

Recall that $\alpha_\lambda(f, g) = 0$ iff $|\langle Wg, T(f)\rangle| \leq R_\lambda(f)$. For significant reduction of the variance of $\alpha_\lambda(f, g)$ we should have $s_\lambda(f) := R_\lambda(f)/\sigma(f) \gg 1$.
This suggests that for significant reduction of noise in $S_\lambda(g)$ we should have $s_\lambda(f) \gg 1$ for all the test functions.

# A more detailed analysis

We can make the analysis more precise. Suppose we have fixed some $a > 0$ and consider every region where $f^{\mathrm{rec}} > a$ as an object found in the reconstruction. Of course, we want to avoid reconstructing false objects.

# A more detailed analysis

We can make the analysis more precise. Suppose we have fixed some $a > 0$ and consider every region where $f^{\mathrm{rec}} > a$ as an object found in the reconstruction. Of course, we want to avoid reconstructing false objects.

Let's assume that we are using characteristic functions of balls as test functions; $f_D$ is the characteristic function of a ball of diameter $D$. Let $\Omega \subset \mathbb{R}^3$ be the region where the functions are reconstructed. The number of disjoint balls in this region is roughly $|\Omega| D^{-3}$.

# A more detailed analysis

Note that $\alpha_\lambda(f_D, G^{\mathrm{noise}}) > a$ iff $\langle WG^{\mathrm{noise}}, T(f_D)\rangle > R_\lambda(f_D) + a\|W^{1/2}T(f_D)\|^2$.

# A more detailed analysis

Note that $\alpha_\lambda(f_D, G^{\text{noise}}) > a$ iff
$\langle WG^{\text{noise}}, T(f_D)\rangle > R_\lambda(f_D) + a\|W^{1/2}T(f_D)\|^2$.
Assuming that $\langle WG^{\text{noise}}, T(f_D)\rangle$ has Gaussian distribution, the number of balls with $\alpha_\lambda(f_D, G^{\text{noise}}) > a$ is approximately

$$\frac{|\Omega|}{2D^3}\operatorname{erfc}\left(\frac{R_\lambda(f_D) + a\|W^{1/2}T(f_D)\|^2}{\sqrt{2}\sigma(f_D)}\right).$$

# A more detailed analysis

Note that $\alpha_\lambda(f_D, G^{\rm noise}) > a$ iff
$\langle WG^{\rm noise}, T(f_D)\rangle > R_\lambda(f_D) + a\|W^{1/2}T(f_D)\|^2$.
Assuming that $\langle WG^{\rm noise}, T(f_D)\rangle$ has Gaussian distribution, the number of balls with $\alpha_\lambda(f_D, G^{\rm noise}) > a$ is approximately

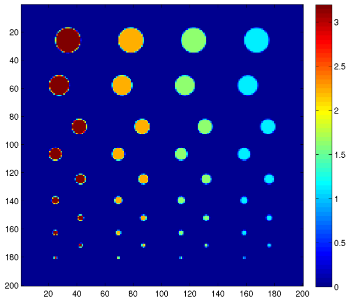
$$\frac{|\Omega|}{2D^3}\,\mathrm{erfc}\left(\frac{R_\lambda(f_D) + a\|W^{1/2}T(f_D)\|^2}{\sqrt{2}\sigma(f_D)}\right).$$

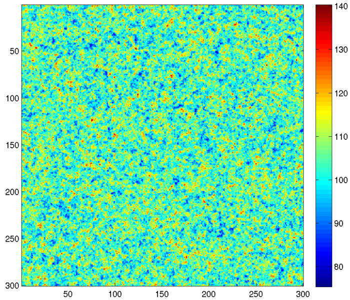
We want this number to be small, which suggests that we should have

$$s_\lambda(f_D) \gtrsim s_{\min}(f_D) := \sqrt{2}\,\mathrm{erfc}^{-1}\left(\frac{2D^3}{|\Omega|}\right) - \frac{a\|W^{1/2}T(f_D)\|^2}{\sigma(f_D)}.$$

# A numerical example

Let's see how this works in practice. We will look at simulated data from a simple phantom containing balls of varying size and contrast.




Above is a section through the phantom (left) and one of the projections (right).

# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

| $D$ | $s_{\min}$ |
|---|---|
| 2.00 | 4.35 |
| 2.40 | 4.12 |
| 2.88 | 3.72 |
| 3.46 | 3.31 |
| 4.15 | 2.61 |
| 4.98 | 1.76 |
| 5.97 | 0.36 |

# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

| $D$ | $s_{\min}$ | $s_{17}$ |
|---|---|---|
| 2.00 | 4.35 | 3.12 |
| 2.40 | 4.12 | 3.05 |
| 2.88 | 3.72 | 2.42 |
| 3.46 | 3.31 | 2.37 |
| 4.15 | 2.61 | 2.09 |
| 4.98 | 1.76 | 1.91 |
| 5.97 | 0.36 | 1.69 |

# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

| $D$ | $s_{\min}$ | $s_{17}$ | $s_{20}$ |
|---|---|---|---|
| 2.00 | 4.35 | 3.12 | 3.68 |
| 2.40 | 4.12 | 3.05 | 3.58 |
| 2.88 | 3.72 | 2.42 | 2.85 |
| 3.46 | 3.31 | 2.37 | 2.78 |
| 4.15 | 2.61 | 2.09 | 2.46 |
| 4.98 | 1.76 | 1.91 | 2.24 |
| 5.97 | 0.36 | 1.69 | 1.99 |

# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

| $D$ | $s_{\min}$ | $s_{17}$ | $s_{20}$ | $s_{24}$ |
|---|---|---|---|---|
| 2.00 | 4.35 | 3.12 | 3.68 | 4.41 |
| 2.40 | 4.12 | 3.05 | 3.58 | 4.30 |
| 2.88 | 3.72 | 2.42 | 2.85 | 3.42 |
| 3.46 | 3.31 | 2.37 | 2.78 | 3.34 |
| 4.15 | 2.61 | 2.09 | 2.46 | 2.95 |
| 4.98 | 1.76 | 1.91 | 2.24 | 2.69 |
| 5.97 | 0.36 | 1.69 | 1.99 | 2.38 |

# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

| $D$ | $s_{\min}$ | $s_{17}$ | $s_{20}$ | $s_{24}$ | $s_{28}$ |
|---|---|---|---|---|---|
| 2.00 | 4.35 | 3.12 | 3.68 | 4.41 | 5.15 |
| 2.40 | 4.12 | 3.05 | 3.58 | 4.30 | 5.02 |
| 2.88 | 3.72 | 2.42 | 2.85 | 3.42 | 3.99 |
| 3.46 | 3.31 | 2.37 | 2.78 | 3.34 | 3.90 |
| 4.15 | 2.61 | 2.09 | 2.46 | 2.95 | 3.44 |
| 4.98 | 1.76 | 1.91 | 2.24 | 2.69 | 3.14 |
| 5.97 | 0.36 | 1.69 | 1.99 | 2.38 | 2.78 |

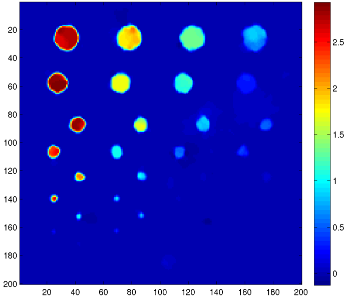
# Numerical example continued

I choose (rather arbitrarily) $a = 0.5$ and compute $s_{\min}(f_D)$ for a range of diameters $D$. These are compared to $s_\lambda(f_D)$ for some different $\lambda$.

| $D$ | $s_{\min}$ | $s_{17}$ | $s_{20}$ | $s_{24}$ | $s_{28}$ |
|---|---|---|---|---|---|
| 2.00 | 4.35 | 3.12 | 3.68 | 4.41 | 5.15 |
| 2.40 | 4.12 | 3.05 | 3.58 | 4.30 | 5.02 |
| 2.88 | 3.72 | 2.42 | 2.85 | 3.42 | 3.99 |
| 3.46 | 3.31 | 2.37 | 2.78 | 3.34 | 3.90 |
| 4.15 | 2.61 | 2.09 | 2.46 | 2.95 | 3.44 |
| 4.98 | 1.76 | 1.91 | 2.24 | 2.69 | 3.14 |
| 5.97 | 0.36 | 1.69 | 1.99 | 2.38 | 2.78 |

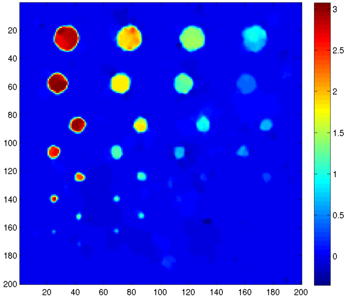
With $\lambda \geq 24$ we should be rather confident not to reconstruct false objects.

# Reconstruction with $\lambda = 28$



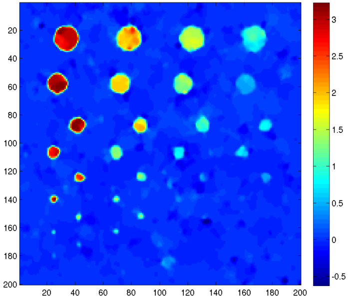

True objects: 20
False objects: 0

# Reconstruction with $\lambda = 24$




True objects: 24
False objects: 1

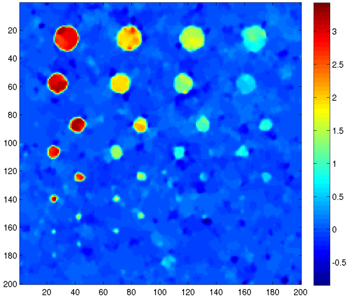
# Reconstruction with $\lambda = 20$




True objects: 25

False objects: 21

# Reconstruction with $\lambda = 17$




True objects: 28
False objects: 137

# Discrepancy principle

One well-known method used to choose regularization parameters is the discrepancy principle. The idea is to regularize so much that the residual norm $\|T(f^{\mathrm{rec}}) - g\|$ agrees with the estimated norm of $\|g^{\mathrm{noise}}\|$.

# Discrepancy principle

One well-known method used to choose regularization parameters is the discrepancy principle. The idea is to regularize so much that the residual norm $\|T(f^{\text{rec}}) - g\|$ agrees with the estimated norm of $\|g^{\text{noise}}\|$.
Here is what the residual norm looks like for our numerical example:

| $\lambda$ | $\|T(f^{\text{rec}}) - g\|/\|g\|$ |
|---|---|
| 28 | 0.999634 |
| 24 | 0.999659 |
| 20 | 0.999704 |
| 17 | 0.999728 |

For the discrepancy principle to be useful, $\|g^{\text{noise}}\|$ would have to be known with 4–5 digits accuracy!

# Example with real data

We apply the same analysis to the real data from the TMV example.

| $D$ | $s_{\min}$ | $s_{30}$ | $s_{60}$ |
|---|---|---|---|
| 2.00 | 4.68 | 2.56 | 5.12 |
| 2.40 | 4.54 | 2.18 | 4.36 |
| 2.88 | 4.38 | 1.95 | 3.90 |
| 3.46 | 4.19 | 1.69 | 3.38 |
| 4.15 | 3.96 | 1.46 | 2.93 |
| 4.98 | 3.65 | 1.28 | 2.55 |
| 5.97 | 3.24 | 1.10 | 2.21 |
| 7.17 | 2.65 | 0.97 | 1.94 |
| 8.60 | 1.75 | 0.87 | 1.74 |
| 10.32 | 0.48 | 0.77 | 1.55 |

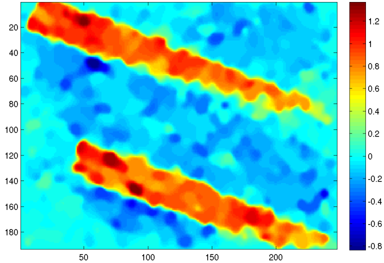
# Example with real data

We apply the same analysis to the real data from the TMV example.

| $D$ | $s_{\min}$ | $s_{30}$ | $s_{60}$ |
|---|---|---|---|
| 2.00 | 4.68 | 2.56 | 5.12 |
| 2.40 | 4.54 | 2.18 | 4.36 |
| 2.88 | 4.38 | 1.95 | 3.90 |
| 3.46 | 4.19 | 1.69 | 3.38 |
| 4.15 | 3.96 | 1.46 | 2.93 |
| 4.98 | 3.65 | 1.28 | 2.55 |
| 5.97 | 3.24 | 1.10 | 2.21 |
| 7.17 | 2.65 | 0.97 | 1.94 |
| 8.60 | 1.75 | 0.87 | 1.74 |
| 10.32 | 0.48 | 0.77 | 1.55 |

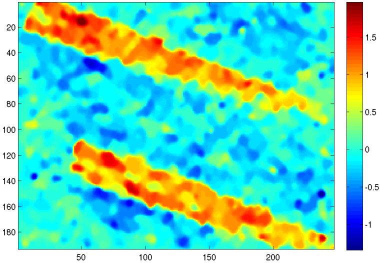
Even with the stronger regularization we should expect some false objects.

# Reconstruction with $\lambda = 60$




| | |
|---|---|
| TMVs: | 2 |
| Other objects: | 17 |

# Reconstruction with $\lambda = 30$




| | |
|---|---|
| TMVs: | 2 |
| Other objects: | 367 |

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$

The variance of $\langle WG^{\mathrm{noise}}, T(f)\rangle$ in an ET tilt series can be estimated, given only $f$, the measured data $g$, the model forward operator $T$, and $W$.

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$

The variance of $\langle WG^{\mathrm{noise}}, T(f)\rangle$ in an ET tilt series can be estimated, given only $f$, the measured data $g$, the model forward operator $T$, and $W$.

This is based on the following circumstances:

1. Noise in different images of the tilt series is uncorrelated.
2. Noise in different parts of the same image is only weakly correlated.
3. The noise is translation invariant.
4. In each image, the SNR is very low.

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$

The variance of $\langle WG^{\mathrm{noise}}, T(f)\rangle$ in an ET tilt series can be estimated, given only $f$, the measured data $g$, the model forward operator $T$, and $W$.

This is based on the following circumstances:

1. Noise in different images of the tilt series is uncorrelated.
2. Noise in different parts of the same image is only weakly correlated.
3. The noise is translation invariant.
4. In each image, the SNR is very low.

Let $Y = Y_1 \oplus \cdots \oplus Y_N$ be the decomposition of the data space into components corresponding to separate images.

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$

The variance of $\langle WG^{\mathrm{noise}}, T(f)\rangle$ in an ET tilt series can be estimated, given only $f$, the measured data $g$, the model forward operator $T$, and $W$.
This is based on the following circumstances:

1. Noise in different images of the tilt series is uncorrelated.
2. Noise in different parts of the same image is only weakly correlated.
3. The noise is translation invariant.
4. In each image, the SNR is very low.

Let $Y = Y_1 \oplus \cdots \oplus Y_N$ be the decomposition of the data space into components corresponding to separate images.
Let $T_i : X \to Y_i$ be the forward operator of the $i$-th projection.
Suppose $W = W_1 \oplus \cdots \oplus W_N$ decomposes into operators on the different images.

# How to estimate $\mathrm{Var}[\langle WG^{\text{noise}}, T(f)\rangle]$

The variance of $\langle WG^{\text{noise}}, T(f)\rangle$ in an ET tilt series can be estimated, given only $f$, the measured data $g$, the model forward operator $T$, and $W$.
This is based on the following circumstances:

1. Noise in different images of the tilt series is uncorrelated.
2. Noise in different parts of the same image is only weakly correlated.
3. The noise is translation invariant.
4. In each image, the SNR is very low.

Let $Y = Y_1 \oplus \cdots \oplus Y_N$ be the decomposition of the data space into components corresponding to separate images.
Let $T_i : X \to Y_i$ be the forward operator of the $i$-th projection.
Suppose $W = W_1 \oplus \cdots \oplus W_N$ decomposes into operators on the different images.
Let $f_1, \ldots f_m$ be (random) translations of $f$.

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$, cont.

By the property 1,

$$\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle] = \sum_{i=1}^{N} \mathrm{Var}[\langle W_i G_i^{\mathrm{noise}}, T_i(f)\rangle].$$

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$, cont.

By the property 1,

$$\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle] = \sum_{i=1}^{N} \mathrm{Var}[\langle W_i G_i^{\mathrm{noise}}, T_i(f)\rangle].$$

By 2 and 3, $\langle Wg_i^{\mathrm{noise}}, T_i(f_1)\rangle, \ldots, \langle Wg_i^{\mathrm{noise}}, T_i(f_m)\rangle$ can be regarded as independent samples of $\langle WG_i^{\mathrm{noise}}, T_i(f)\rangle$, so the variance of $\langle WG_i^{\mathrm{noise}}, T_i(f)\rangle$ can be estimated as a sample variance:

$$\mathrm{Var}[\langle WG_i^{\mathrm{noise}}, T_i(f)\rangle] \approx \frac{1}{m-1}\sum_{j=1}^{m}\left(\langle Wg_i^{\mathrm{noise}}, T_i(f_j)\rangle - \frac{1}{m}\sum_{k=1}^{m}\langle Wg_i^{\mathrm{noise}}, T_i(f_k)\rangle\right)^2.$$

# How to estimate $\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle]$, cont.

By the property 1,

$$\mathrm{Var}[\langle WG^{\mathrm{noise}}, T(f)\rangle] = \sum_{i=1}^{N} \mathrm{Var}[\langle W_i G_i^{\mathrm{noise}}, T_i(f)\rangle].$$

By 2 and 3, $\langle Wg_i^{\mathrm{noise}}, T_i(f_1)\rangle, \ldots, \langle Wg_i^{\mathrm{noise}}, T_i(f_m)\rangle$ can be regarded as independent samples of $\langle WG_i^{\mathrm{noise}}, T_i(f)\rangle$, so the variance of $\langle WG_i^{\mathrm{noise}}, T_i(f)\rangle$ can be estimated as a sample variance:

$$\mathrm{Var}[\langle WG_i^{\mathrm{noise}}, T_i(f)\rangle] \approx \frac{1}{m-1} \sum_{j=1}^{m} \left( \langle Wg_i^{\mathrm{noise}}, T_i(f_j)\rangle - \frac{1}{m} \sum_{k=1}^{m} \langle Wg_i^{\mathrm{noise}}, T_i(f_k)\rangle \right)^2 .$$

Finally, by 4, we can replace $g^{\mathrm{noise}}$ by $g$ in the last expression.