Stochastic Optimization from the perspective of dynamical systems

Emmanouil Daskalakis
Felix J. Herrmann
Rachel Kuske
Mengmeng Yang

SLIM
University of British Columbia

Motivation: least-squares migration

Consider

$$\min_{\mathbf{x}} ||\mathbf{x}||_1$$

s.t.

$$\sum_{i=1}^{n_s} ||\mathbf{J}_i[\mathbf{m}_0, \mathbf{q}_i]\mathbf{C}^*\mathbf{x} - \mathbf{b}_i||_2 \leq \sigma$$

- $\mathbf{x}$ is the vector of Curvelet coefficients,
- $\mathbf{J}_i$ is the Born modelling operator,
- $\mathbf{m}_0$ is the background model for the velocity,
- $\mathbf{b}_i$ is the vectorized reflection of the i-th shot,
- $\mathbf{C}^*$ is the transpose of the curvelet transform,
- σ is the tolerance for noise.

Motivation: least-squares migration

Consider

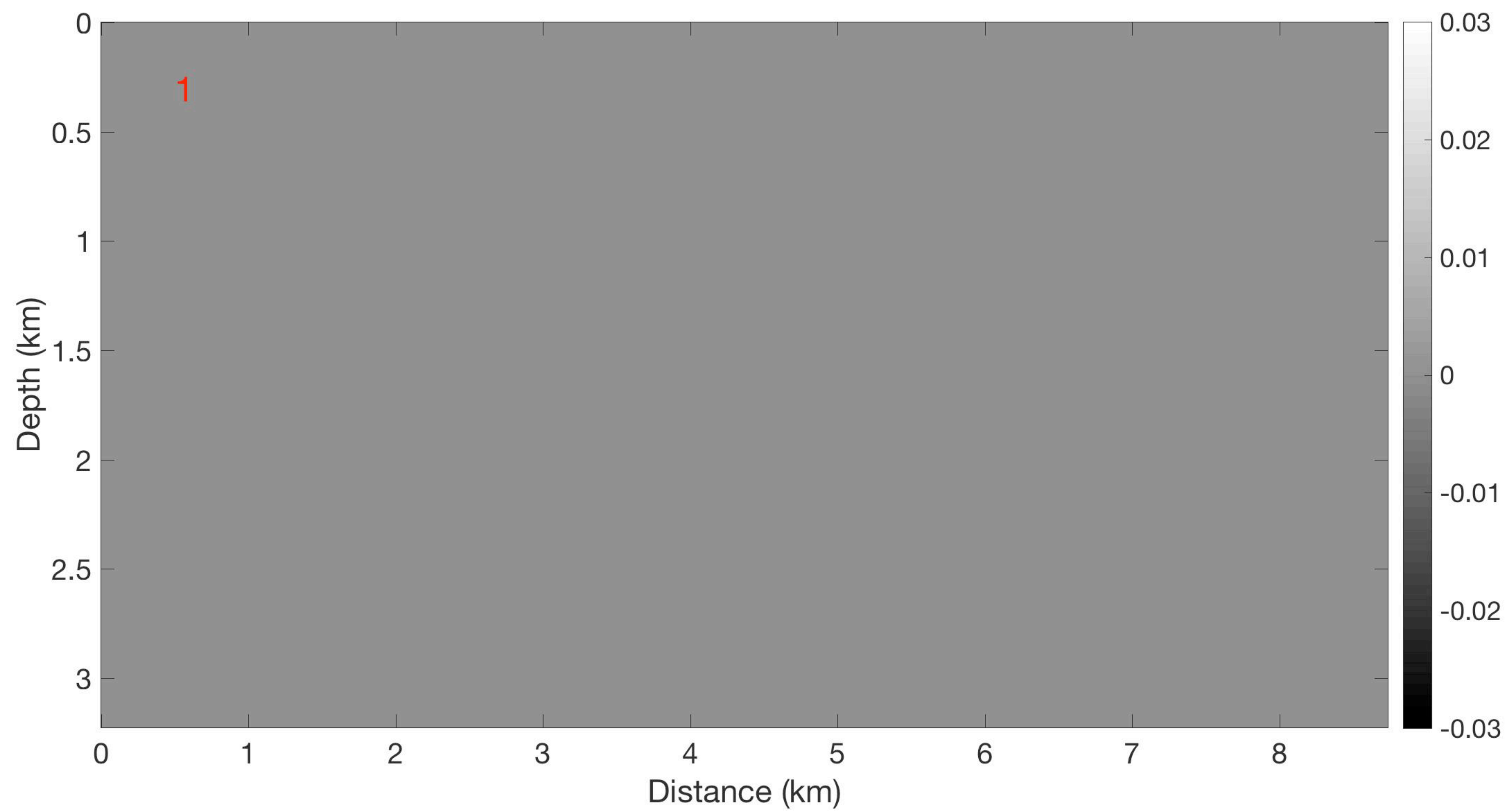
$$\min_{\mathbf{x}} ||\mathbf{x}||_1$$

s.t.

$$\sum_{i=1}^{n_s} ||\mathbf{J}_i[\mathbf{m}_0, \mathbf{q}_i]\mathbf{C}^*\mathbf{x} - \mathbf{b}_i||_2 \leq \sigma$$

- $\mathbf{x}$ is the vector of Curvelet coefficients,
- $\mathbf{J}_i$ is the Born modelling operator,
- $\mathbf{m}_0$ is the background model for the velocity,
- $\mathbf{b}_i$ is the vectorized reflection of the i-th shot,
- $\mathbf{C}^*$ is the transpose of the curvelet transform,
- σ is the tolerance for noise.

Solution by LB:



Motivation: least-squares migration

We consider the following optimization problem:

$$\min_{\mathbf{x}} ||\mathbf{x}||_1$$

s.t.

$$\sum_{i=1}^{n_s} ||\mathbf{J}_i[\mathbf{m}_0, \mathbf{q}_i]\mathbf{C}^*\mathbf{x} - \mathbf{b}_i||_2 \leq \sigma$$

- $\mathbf{x}$ is the vector of Curvelet coefficients,
- $\mathbf{J}_i$ is the Born modelling operator,
- $\mathbf{m}_0$ is the background model for the velocity,
- $\mathbf{b}_i$ is the vectorized reflection of the i-th shot,
- $\mathbf{C}^*$ is the transpose of the curvelet transform,
- σ is the tolerance for noise.

Stochastic optimization problems:

1. l_1-minimization problem (consistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad Ax = b$$

2. BPDN problem (inconsistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||Ax - b||_2 \leq \sigma$$

Theory (for compressive sensing problems)

1. l_1-minimization problem (consistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad Ax = b$$

2. BPDN problem (inconsistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||Ax - b||_2 \leq \sigma$$

Theory (for compressive sensing problems)

1. l_1-minimization problem (consistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad Ax = b$$

2. BPDN problem (inconsistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||Ax - b||_2 \leq \sigma$$

- Donoho, D. L. (2006), "For most large underdetermined systems of linear equations the minimal ℓ1-norm solution is also the sparsest solution." Comm. Pure Appl. Math., 59: 797–829. doi:10.1002/cpa.20132

Theory (for compressive sensing problems)

1. l_1-minimization problem (consistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad Ax = b$$

- Donoho, D. L. (2006), "For most large underdetermined systems of linear equations the minimal ℓ1-norm solution is also the sparsest solution." Comm. Pure Appl. Math., 59: 797–829. doi:10.1002/cpa.20132

2. BPDN problem (inconsistent)

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||Ax - b||_2 \leq \sigma$$

- "Atomic Decomposition by Basis Pursuit" ,Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, SIAM Review 2001 43:1, 129-159

- Candès, E. J., Romberg, J. K. and Tao, T. (2006), "Stable signal recovery from incomplete and inaccurate measurements." Comm. Pure Appl. Math., 59: 1207–1223.

Toy problem

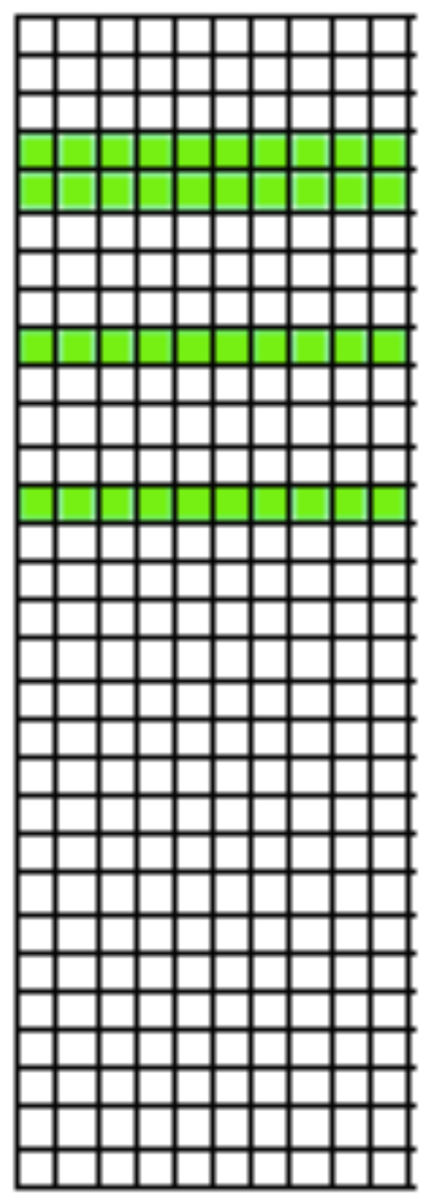
$A \in \mathbb{R}^{20000 \times 1000}$

$b \in \mathbb{R}^{20000}$

Noisy data vector

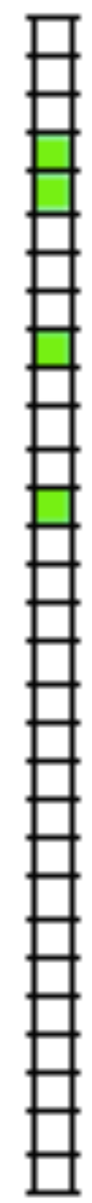
$x \in \mathbb{R}^{1000}$

Sparse vector



*

=



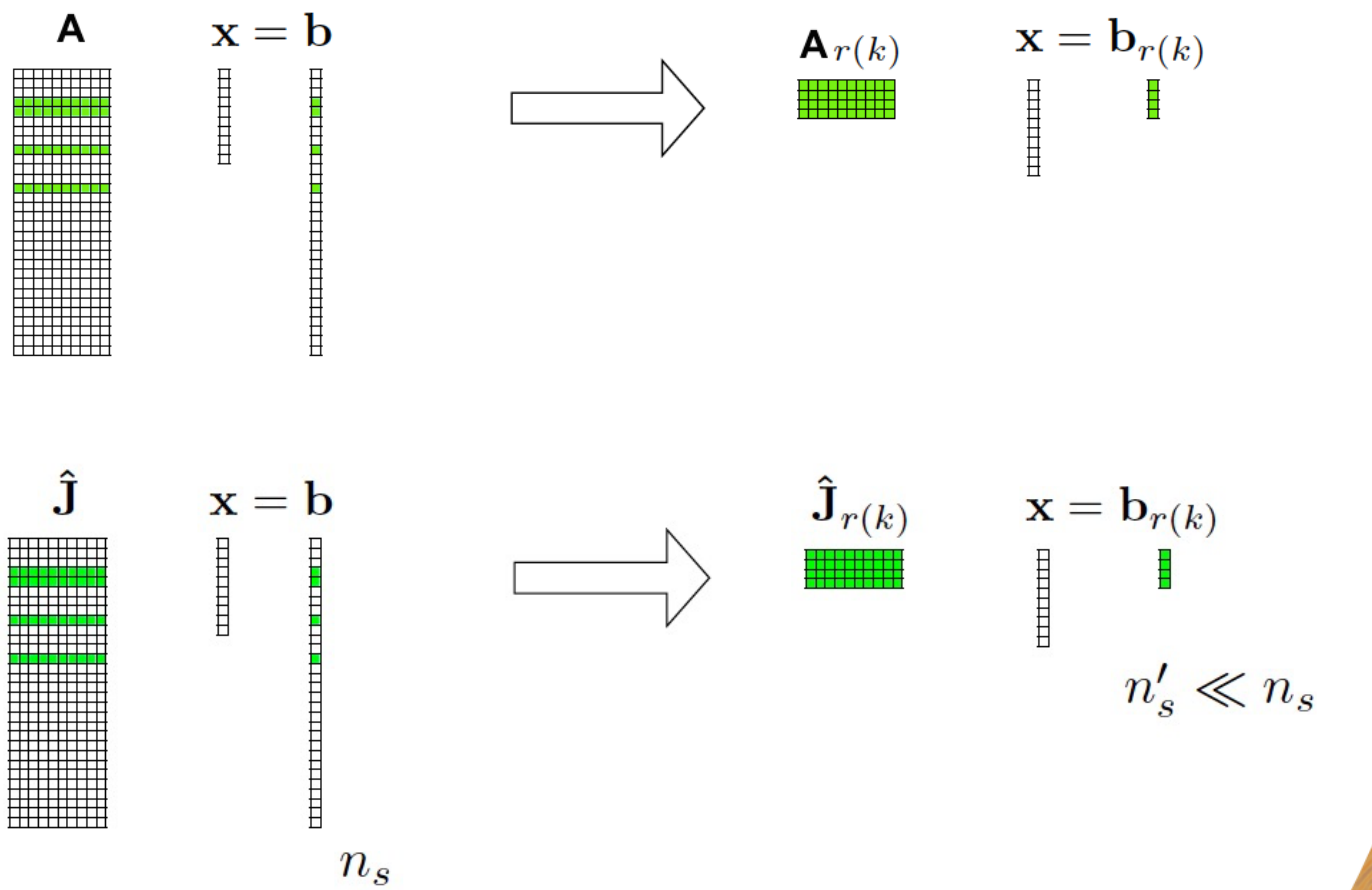
Subsampling

BPDN problem

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||Ax - b||_2 \leq \sigma$$

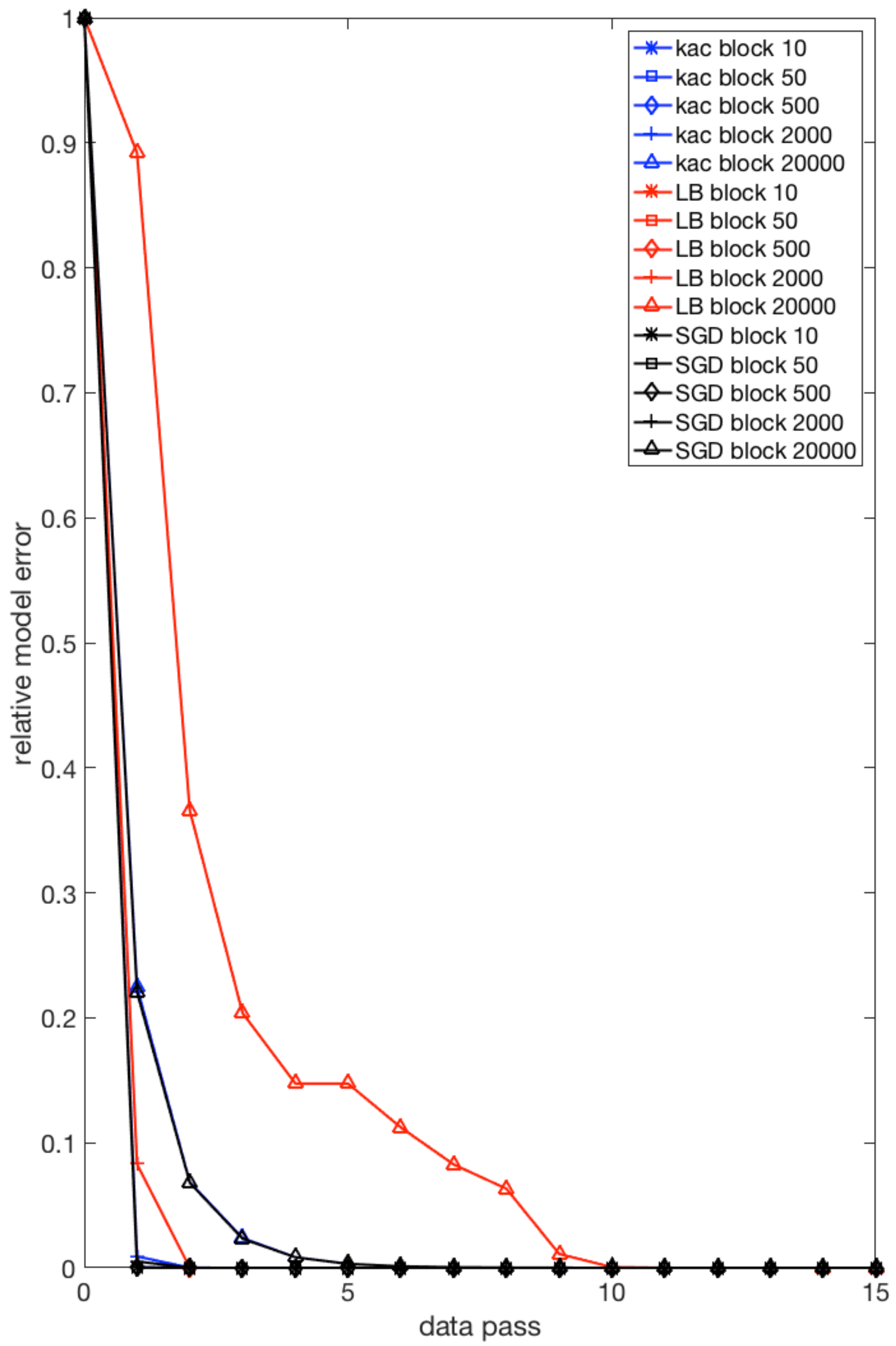
LSRTM problem

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad \sum_{i=1}^{n_s} ||\mathbf{J}_i[\mathbf{m}_0, \mathbf{q}_i]\mathbf{C}^*\mathbf{x} - \mathbf{b}_i||_2 \leq \sigma$$

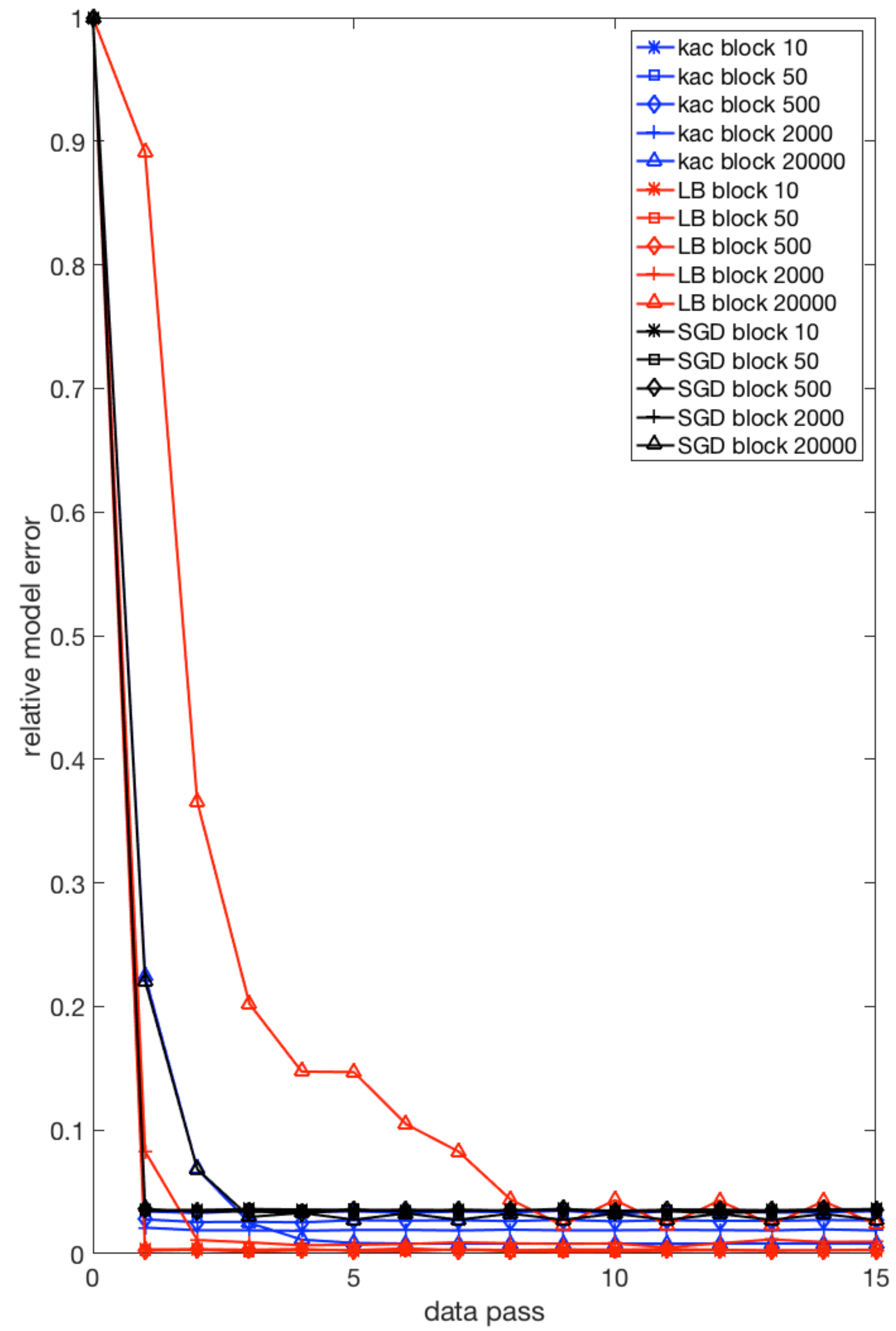


Subsampling

Well condition matrix, noise free data

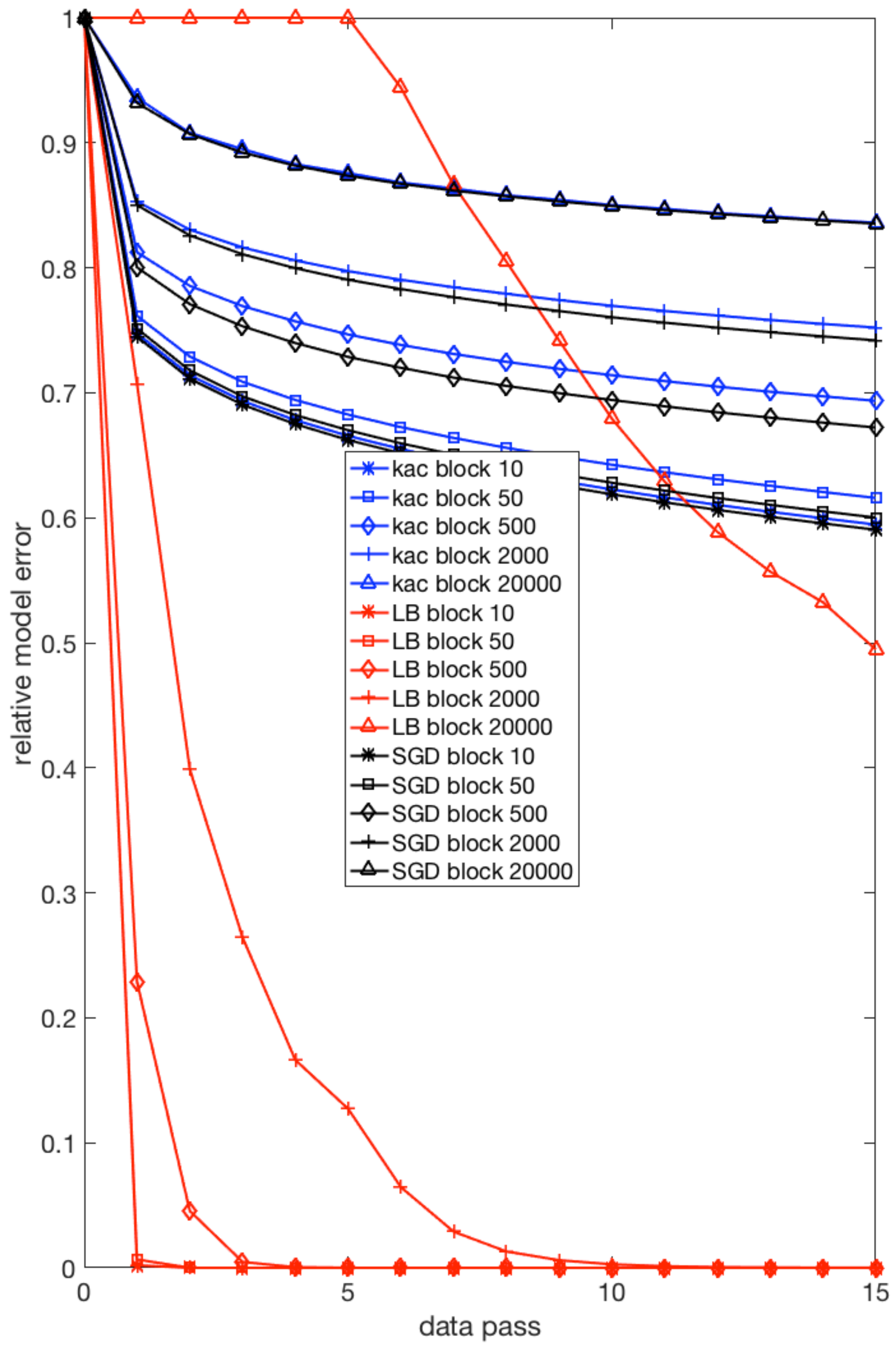


Well condition matrix, noisy data

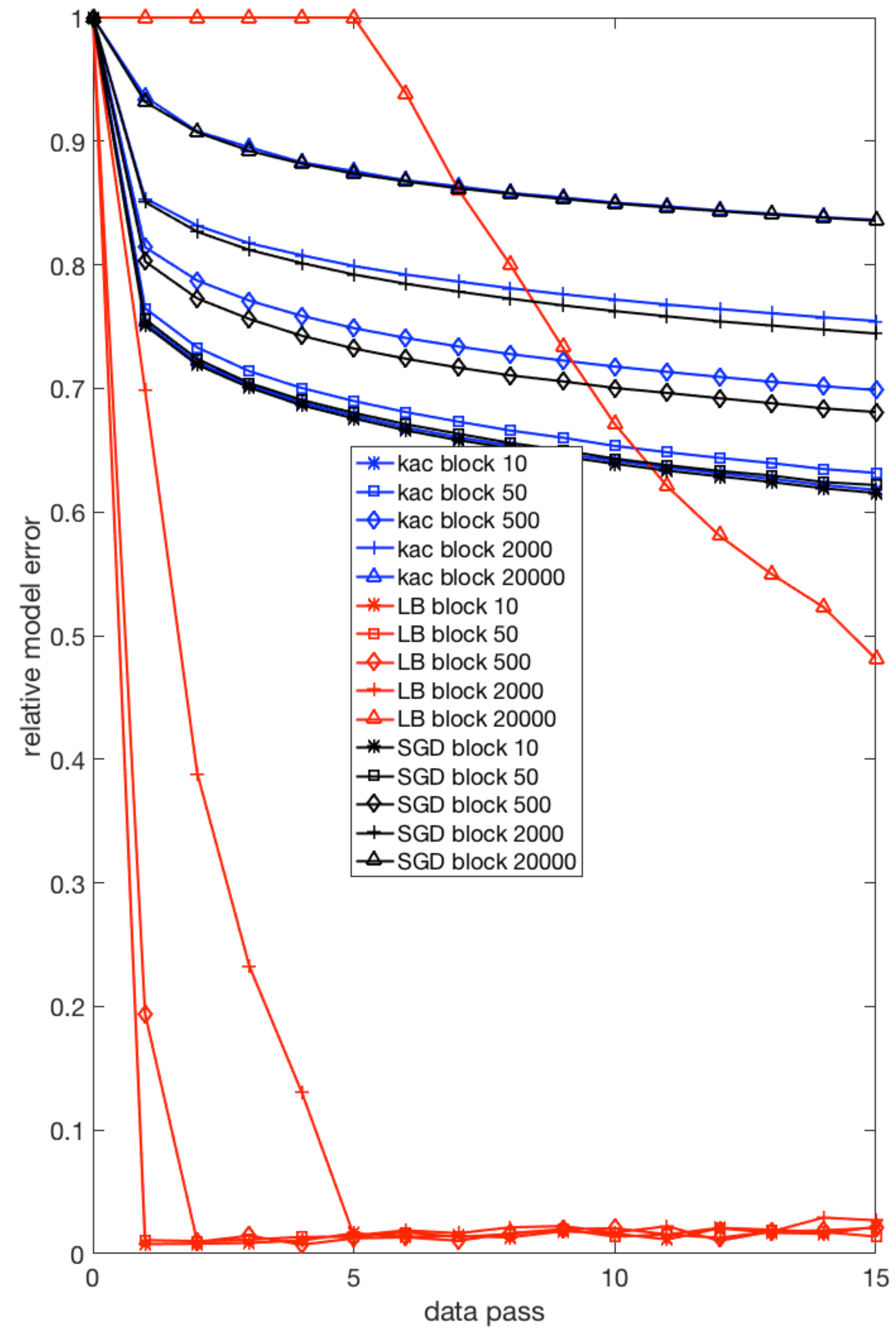


Subsampling

Ill condition matrix, noise free data



Ill condition matrix, noisy data



Least-squares migration problem

LB method

$$\begin{aligned} z_{k+1} &= z_k - t_k A_k^T (A_k x_k - b_k) \\ x_{k+1} &= S_\lambda(z_{k+1}), \end{aligned}$$

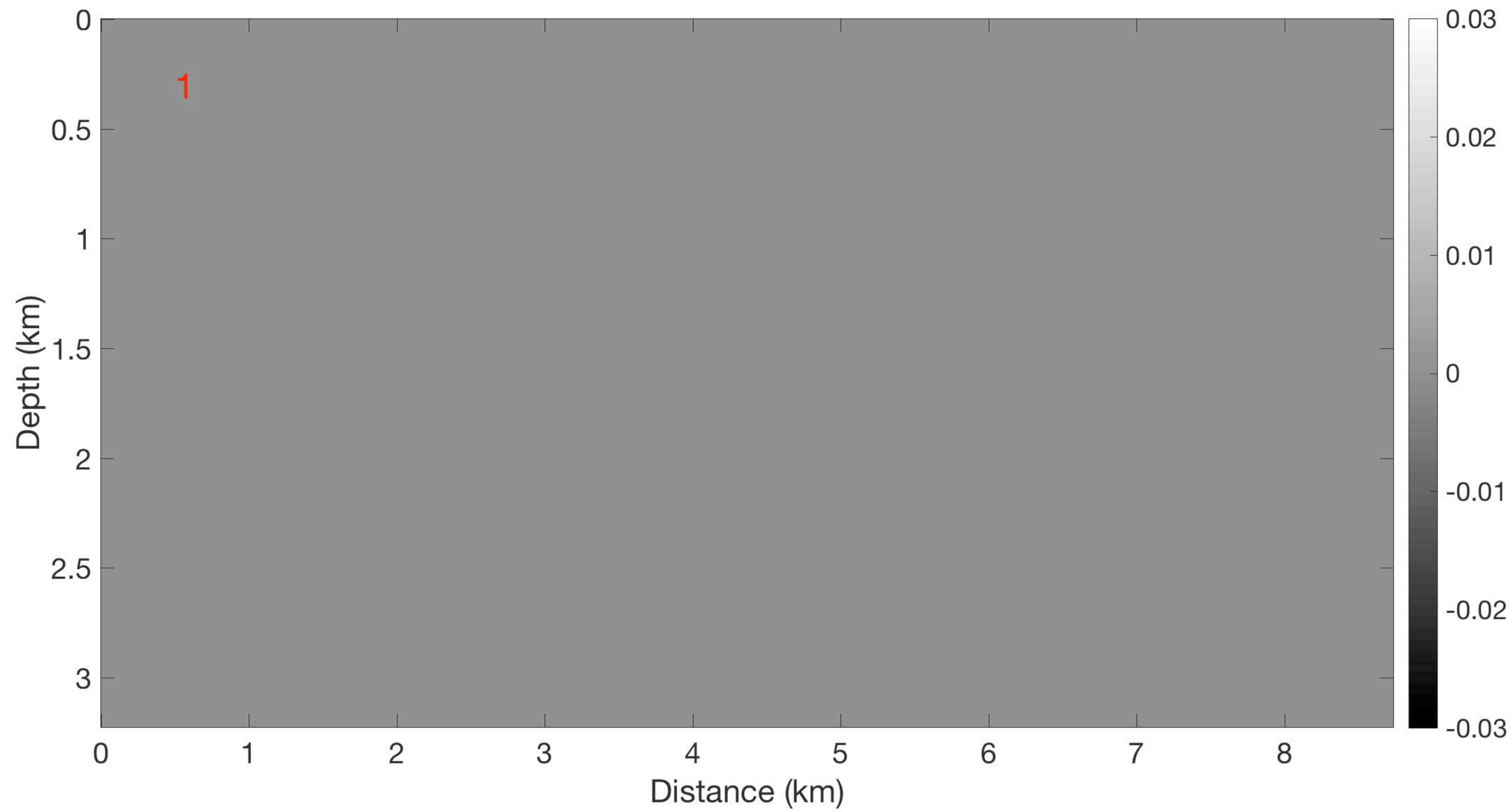
where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\text{sign}(z_k)$$

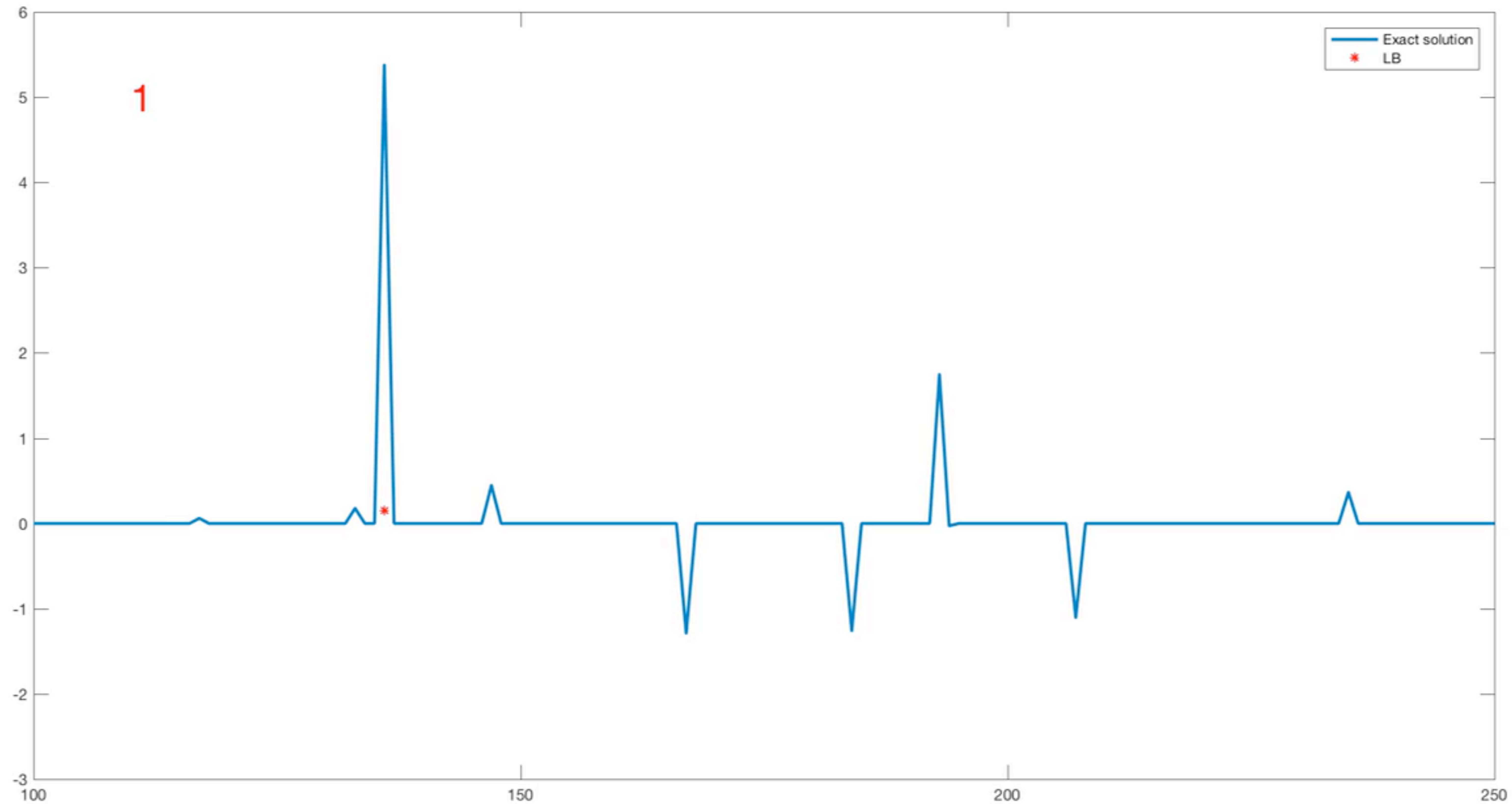
and

$$t_k = \frac{||A_k x_k - b_k||_2^2}{||A_k^T (A_k x_k - b_k)||_2^2}$$

LSRTM using LB



Dealing w/ noisy data



Dealing w/ noisy data

LB method

$$
\begin{aligned}
z_{k+1} &= z_k - t_k A_k^T (A_k x_k - b_k) \\
x_{k+1} &= S_\lambda(z_{k+1}),
\end{aligned}
$$

where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\mathrm{sign}(z_k)$$

and

$$t_k = \frac{||A_k x_k - b_k||_2^2}{||A_k^T (A_k x_k - b_k)||_2^2}$$

LB method w/ L2 ball projection

$$
\begin{aligned}
z_{k+1} &= z_k - t_k A_k^T \Pi_\sigma (A_k x_k - b_k) \\
x_{k+1} &= S_\lambda(z_{k+1}),
\end{aligned}
$$

where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\mathrm{sign}(z_k)$$

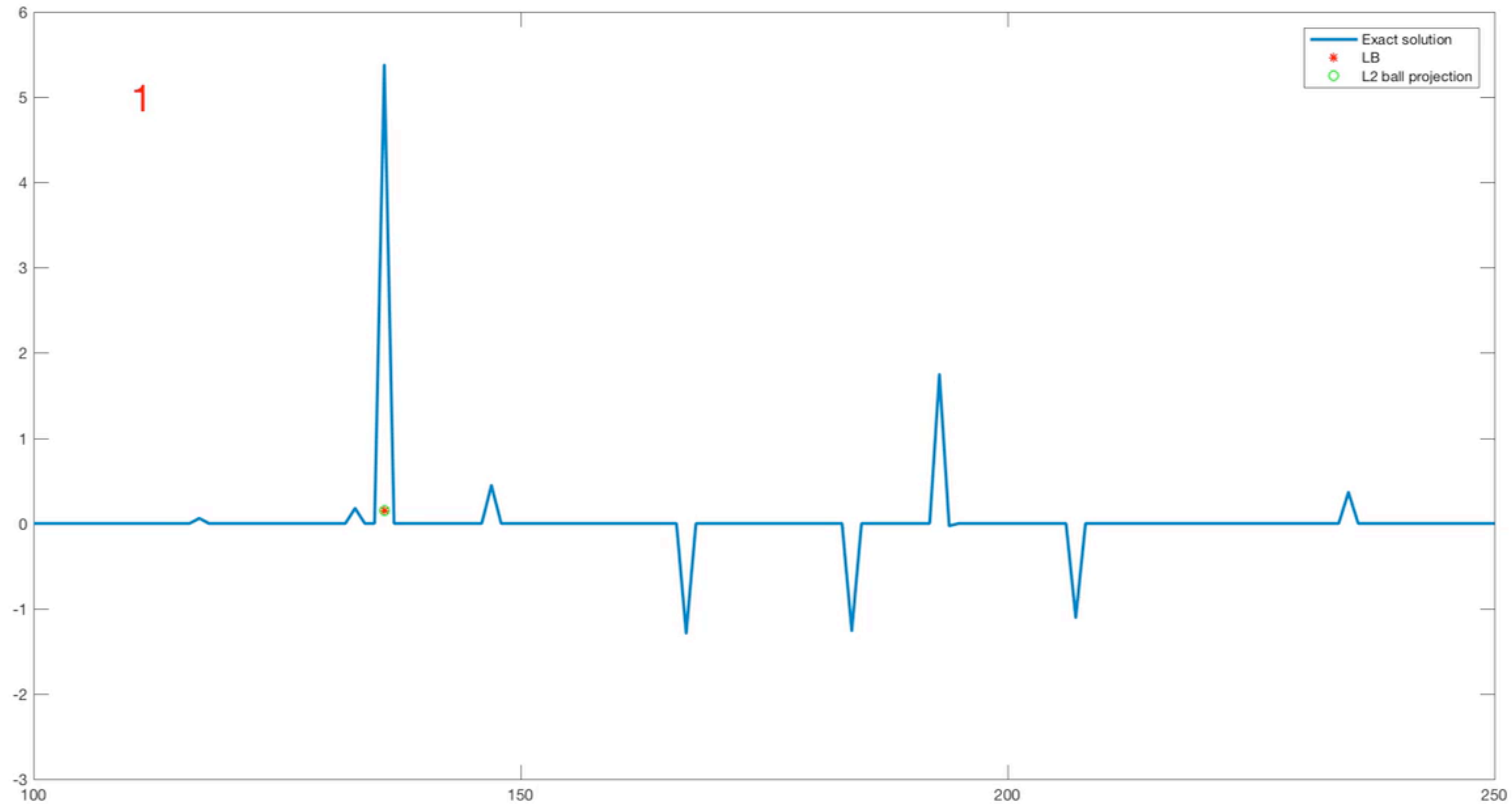
and $$t_k = \frac{||A_k x_k - b_k||_2^2}{||A_k^T (A_k x_k - b_k)||_2^2}$$

Note:

$$\Pi_\sigma(x) = \max\{1 - \frac{\sigma}{||x||_2}\}(x)$$

the projection on to l_2-norm ball

Dealing w/ noisy data



SLIM

Dealing w/ noisy data

LB method

$$
\begin{aligned}
z_{k+1} &= z_k - t_k A_k^T (A_k x_k - b_k) \\
x_{k+1} &= S_\lambda(z_{k+1}),
\end{aligned}
$$

where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\mathrm{sign}(z_k)$$

and

$$t_k = \frac{||A_k x_k - b_k||_2^2}{||A_k^T (A_k x_k - b_k)||_2^2}$$

LB method w/ weighted increment

$$
\begin{aligned}
z_{k+1} &= z_k - \tau_k \odot A_k^T (A_k x_k - b_k) \\
x_{k+1} &= S_\lambda(z_{k+1}),
\end{aligned}
$$

where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\mathrm{sign}(z_k)$$

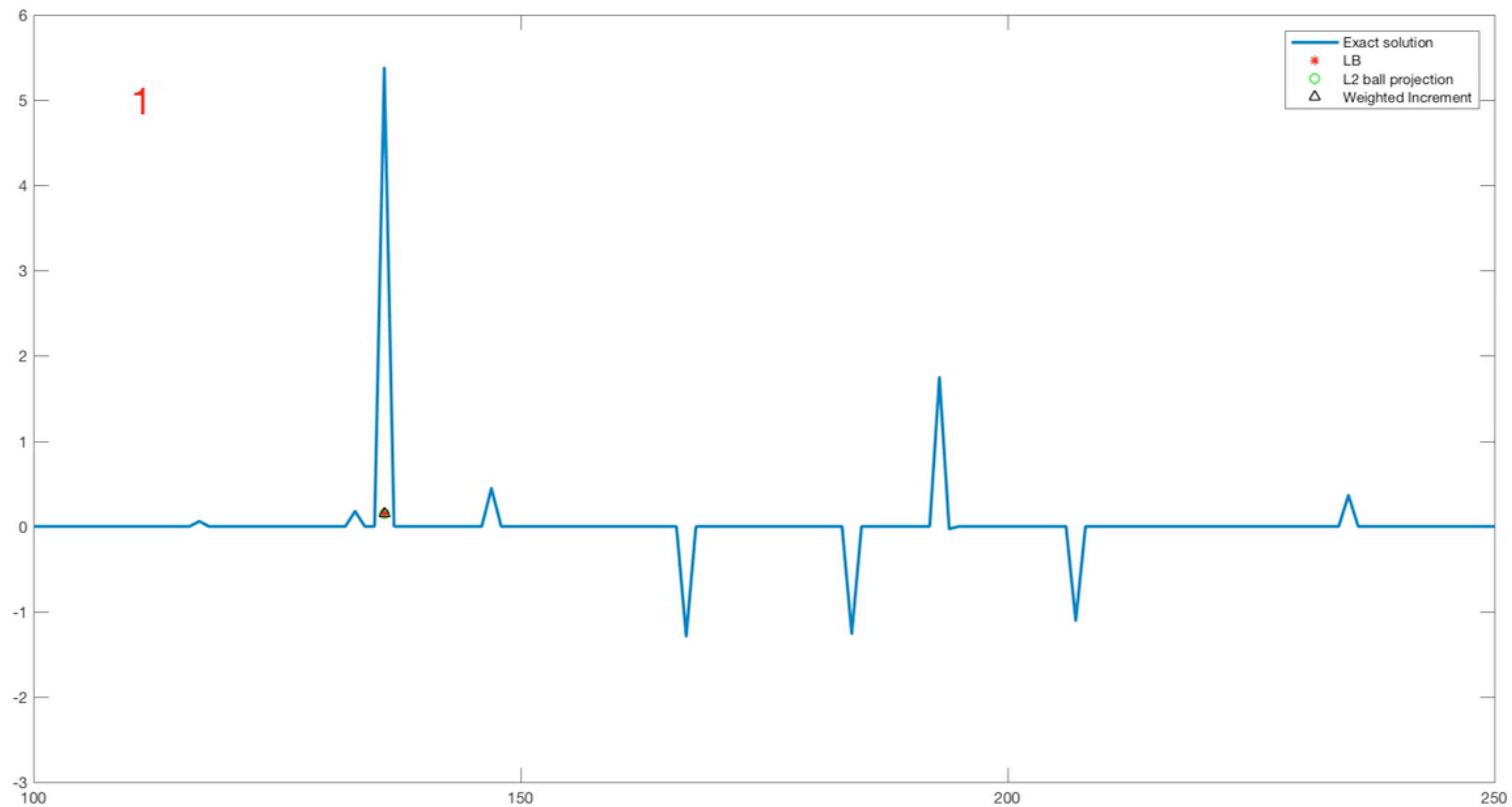
and

$$\tau_k^i = t_k \frac{|\sum_{j=1}^{k} \mathrm{sign}([A_j^T (A_j x_j - b_j)]_i)|}{k}$$

with

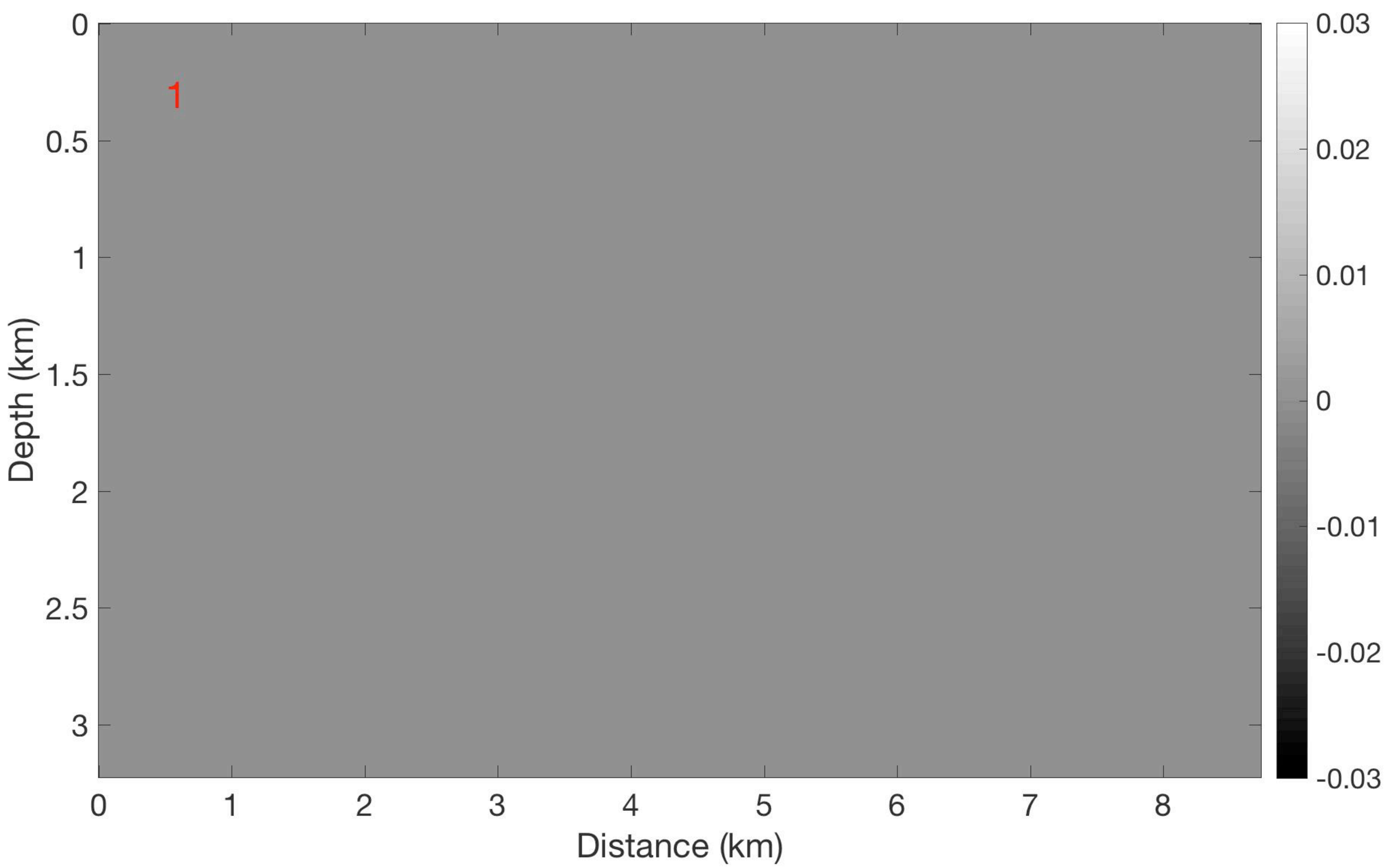
$$t_k = \frac{||A_k x_k - b_k||_2^2}{||A_k^T (A_k x_k - b_k)||_2^2}$$

Dealing w/ noisy data

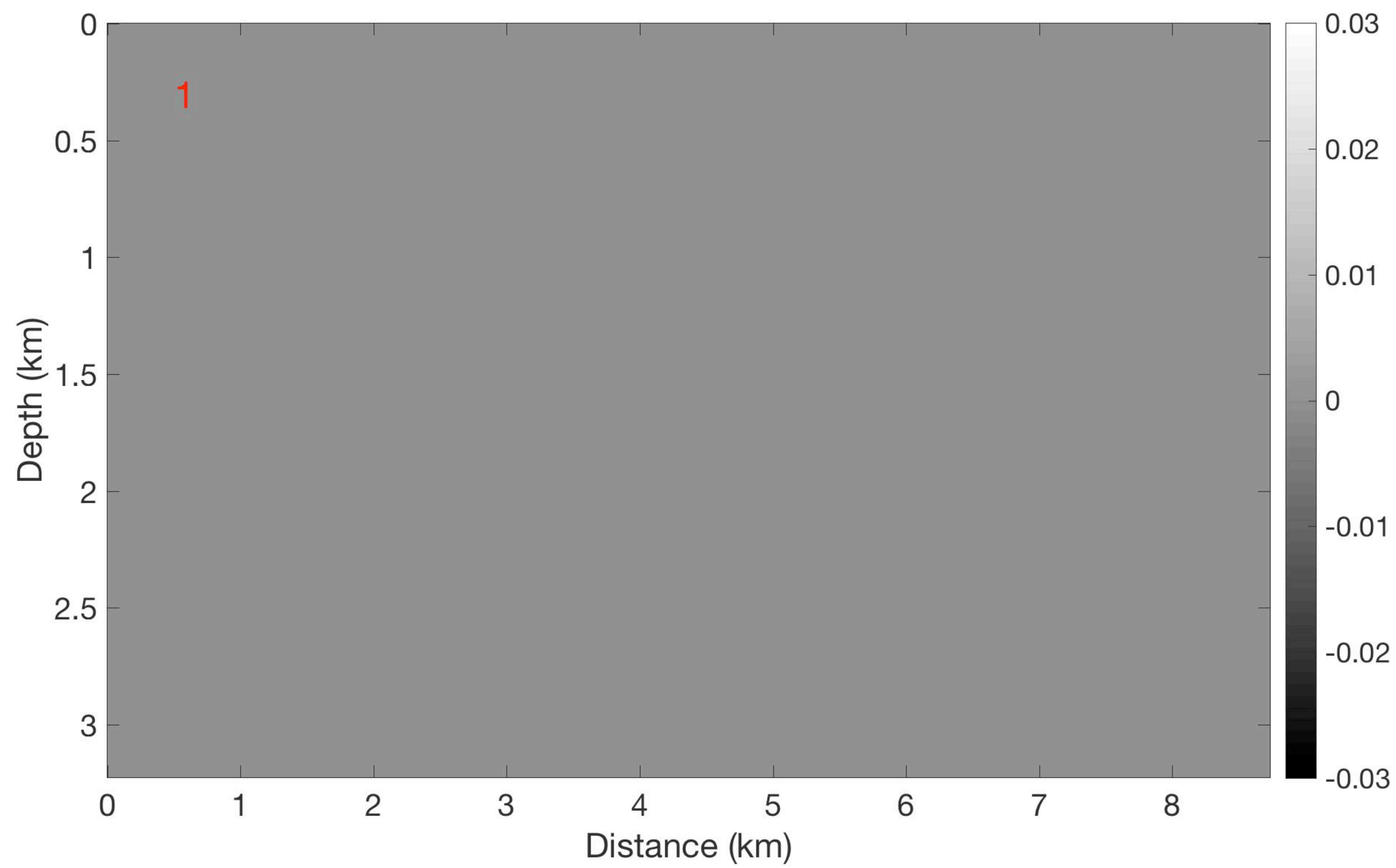


Dealing w/ noisy data

LB method



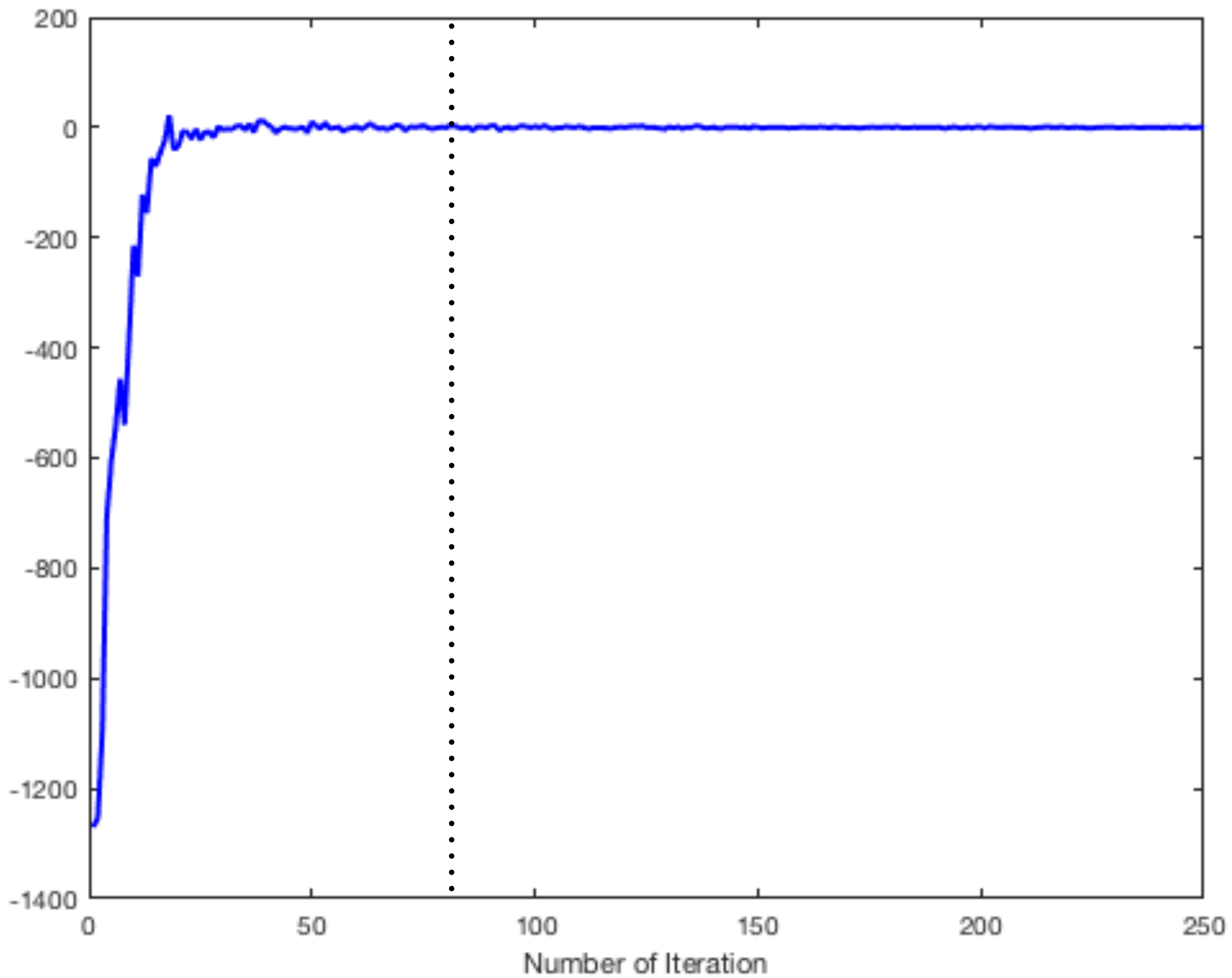
LB method w/ weighted increment



Intuition: gradient entry for weighted increments

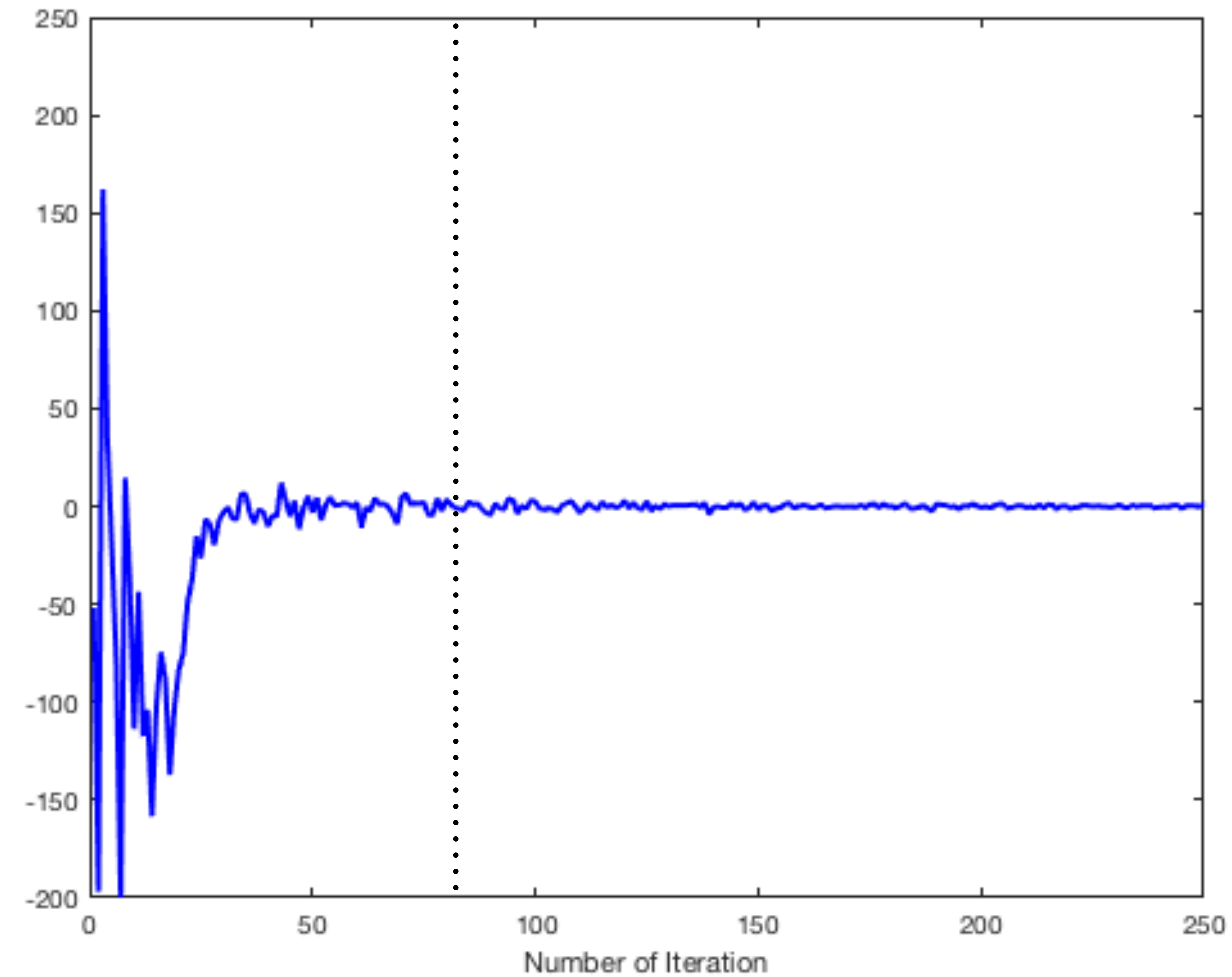
(for consistent problems)

$[A_k^T(A_kx_k - b_k)]_{136}$



Largest entry of the exact solution x^*

$[A_k^T(A_kx_k - b_k)]_{147}$

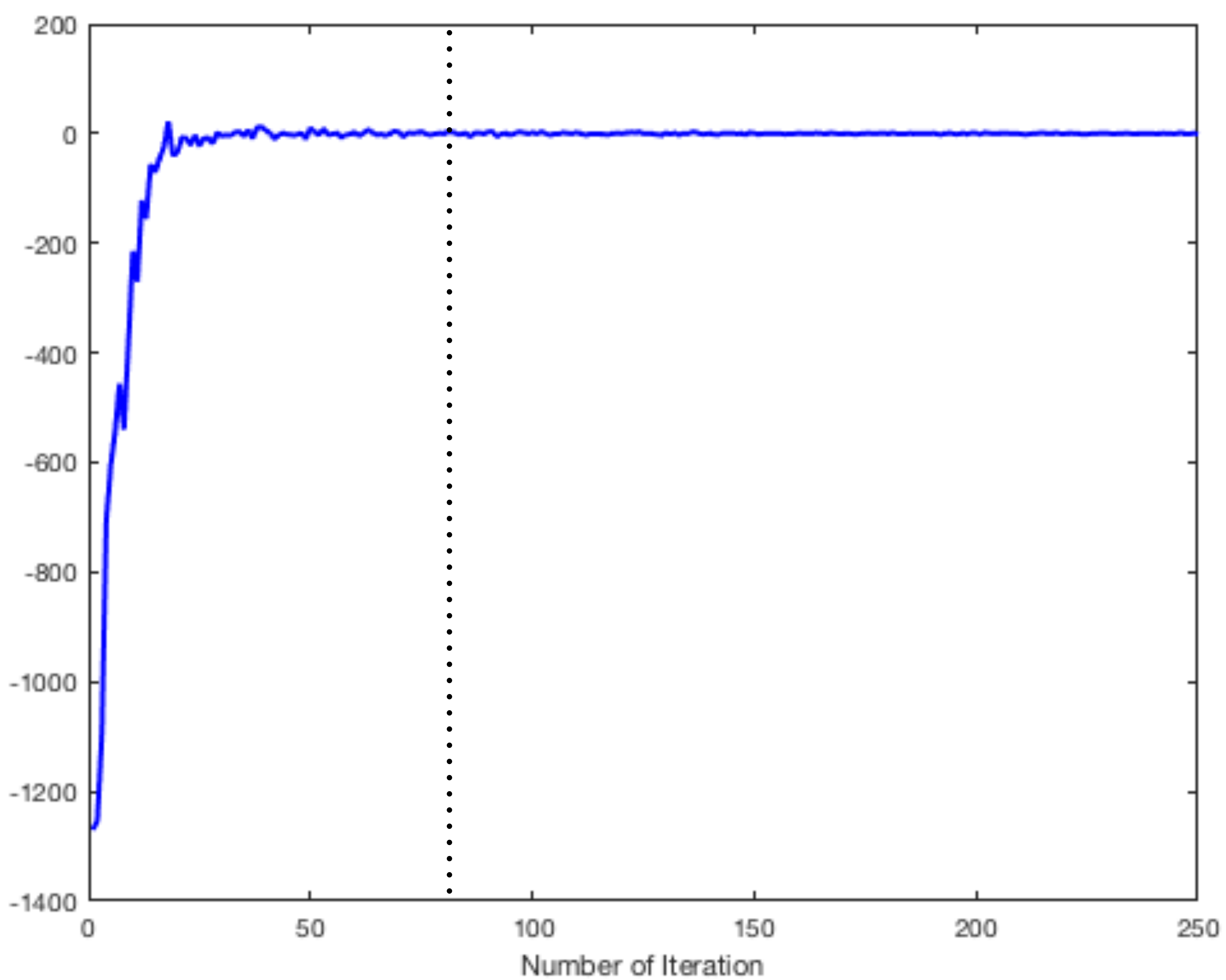


One of the small entries of the exact solution x^*

Intuition: gradient entry for weighted increments

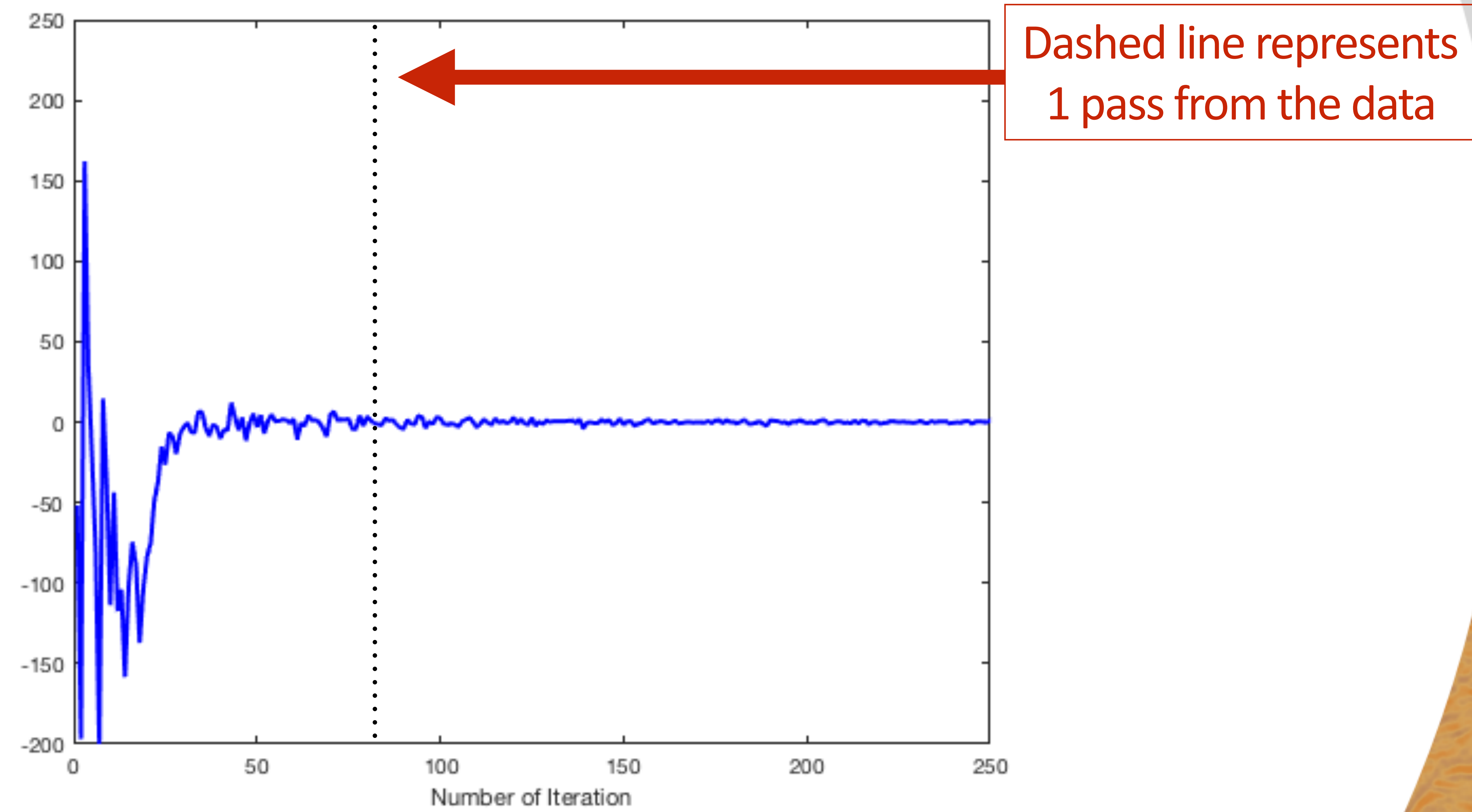
(for consistent problems)

$$[A_k^T(A_kx_k - b_k)]_{136}$$



Largest entry of the exact solution x^*

$$[A_k^T(A_kx_k - b_k)]_{147}$$

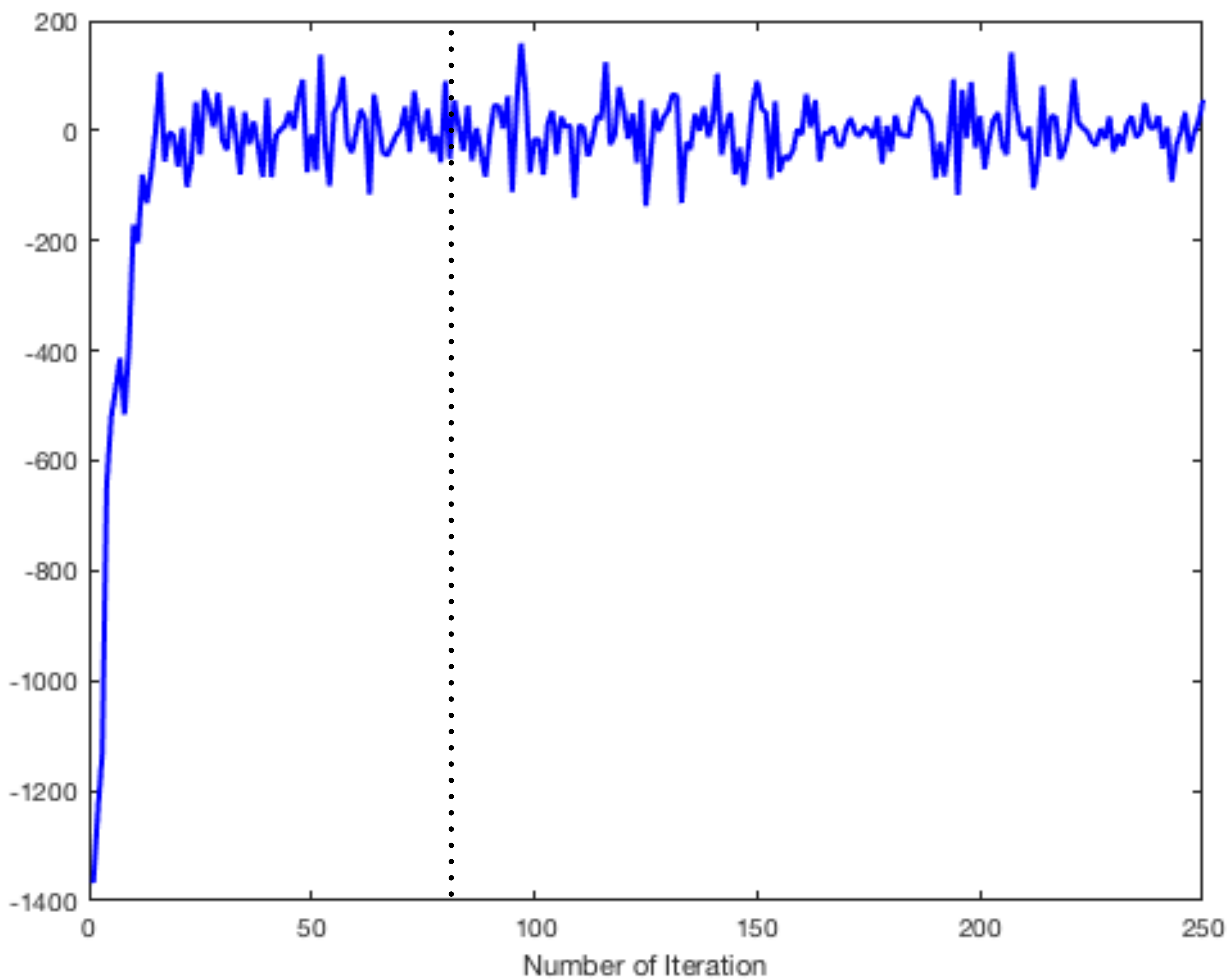


One of the small entries of the exact solution x^*

Intuition: gradient entry for weighted increments

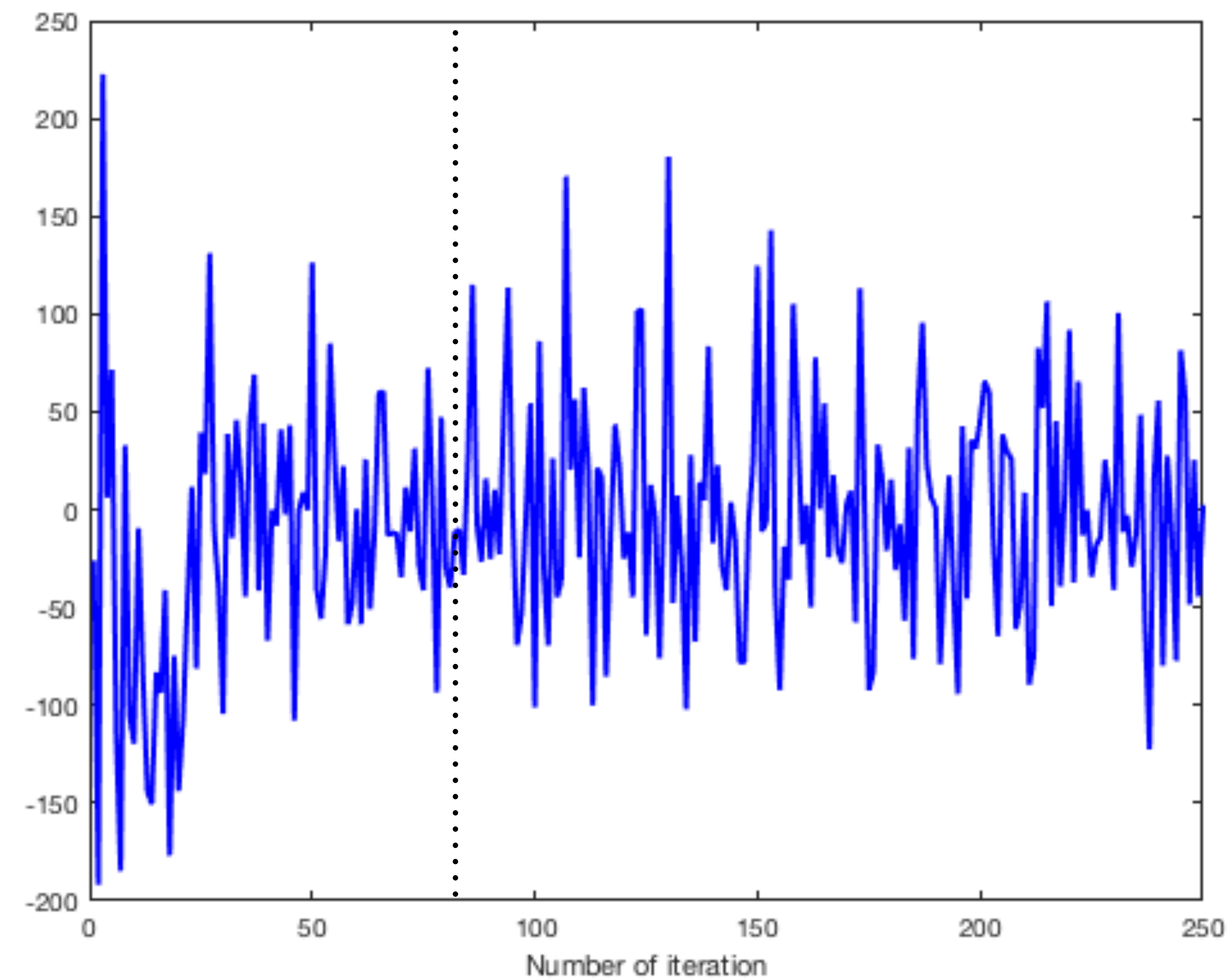
(for inconsistent problem)

$[A_k^T(A_k x_k - b_k)]_{136}$



Largest entry of the exact solution x^*

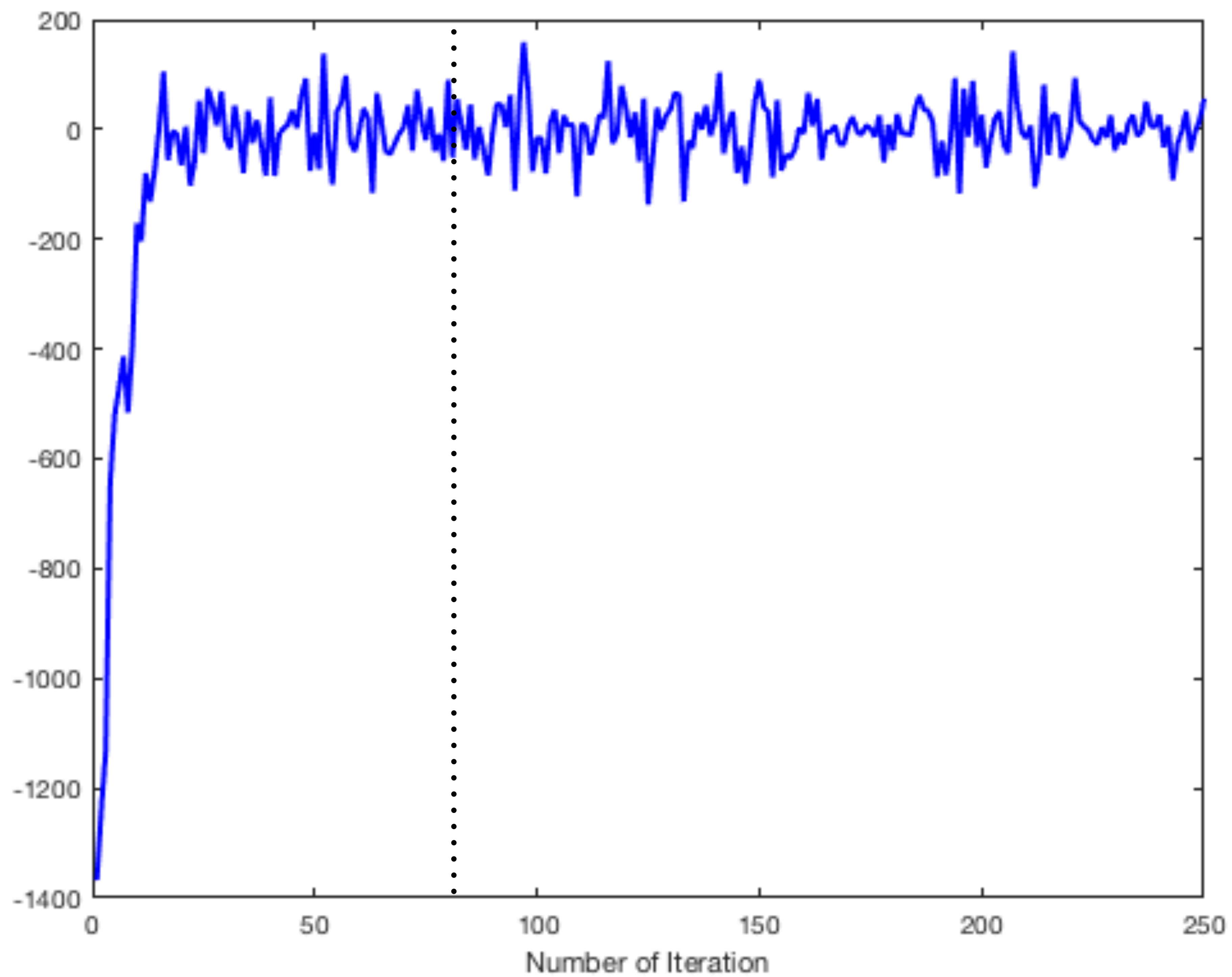
$[A_k^T(A_k x_k - b_k)]_{147}$



One of the small entries of the exact solution x^*

Intuition: Behaviour of the new weighted increment

$$[A_k^T(A_k x_k - b_k)]_{136}$$

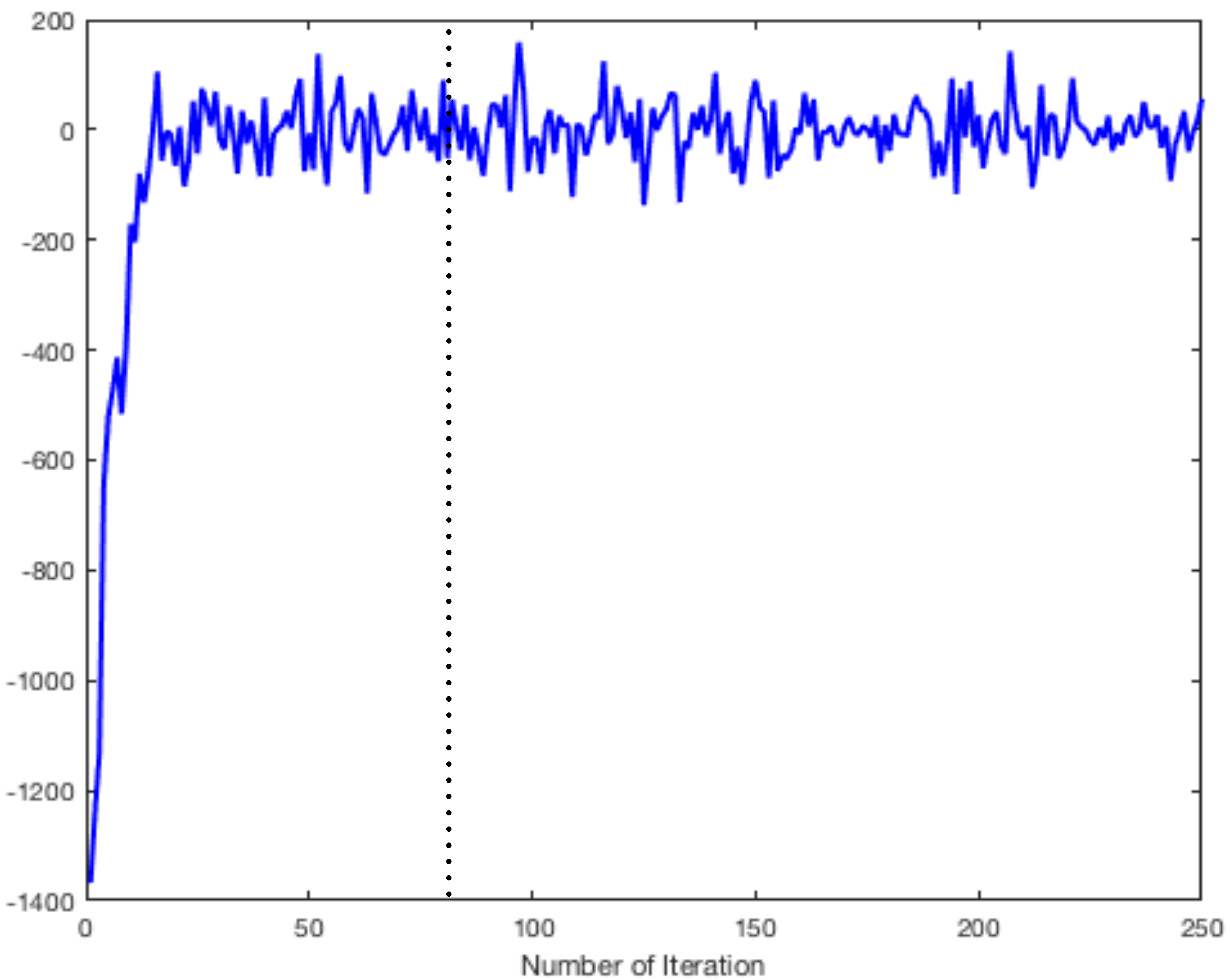


Weighted increment

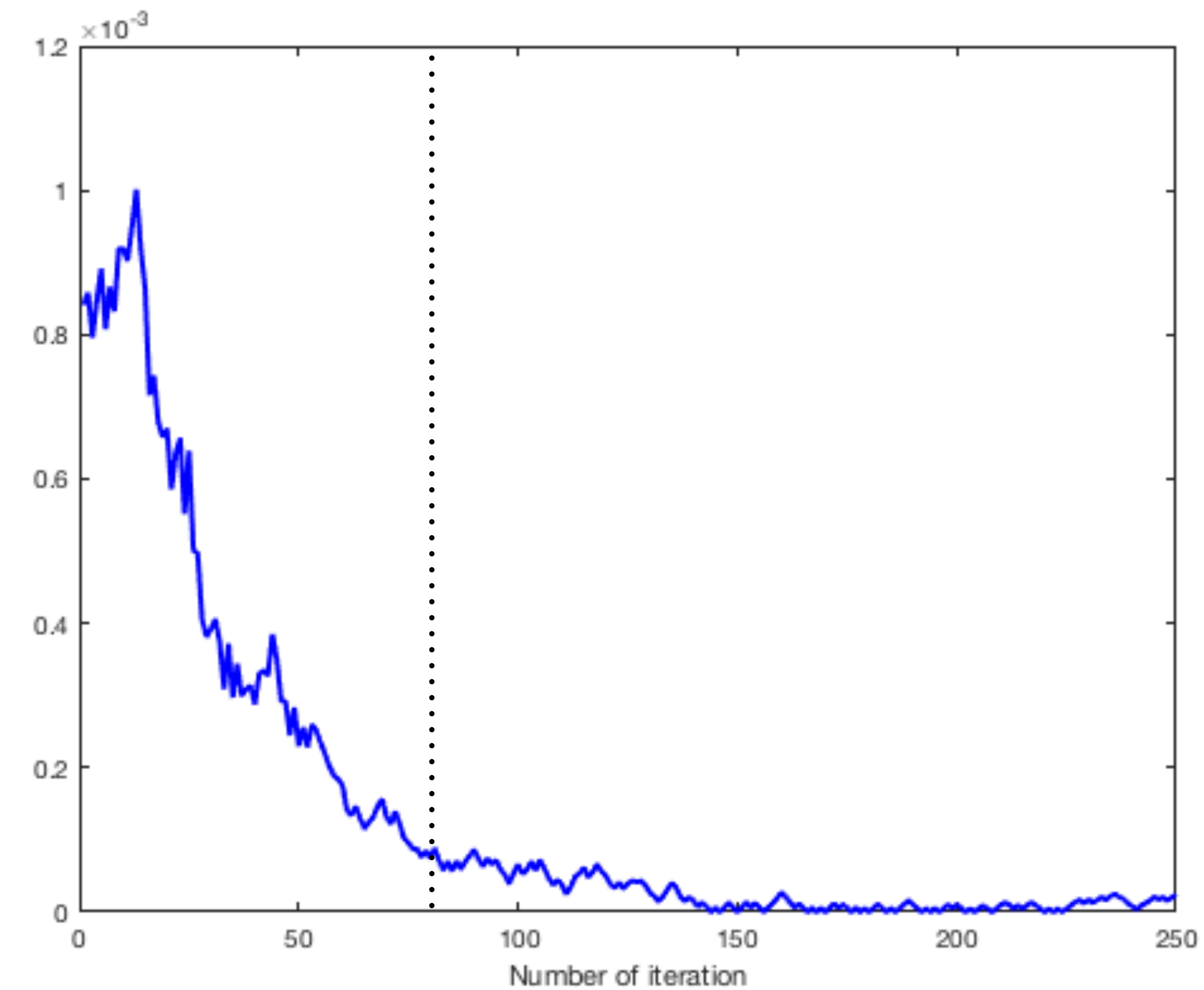
$$\tau_k^i = t_k \frac{|\sum_{j=1}^{k} \mathrm{sign}([A_j^T(A_j x_j - b_j)]_i)|}{k}$$

Intuition: Behaviour of the new weighted increment

$$[A_k^T(A_k x_k - b_k)]_{136}$$

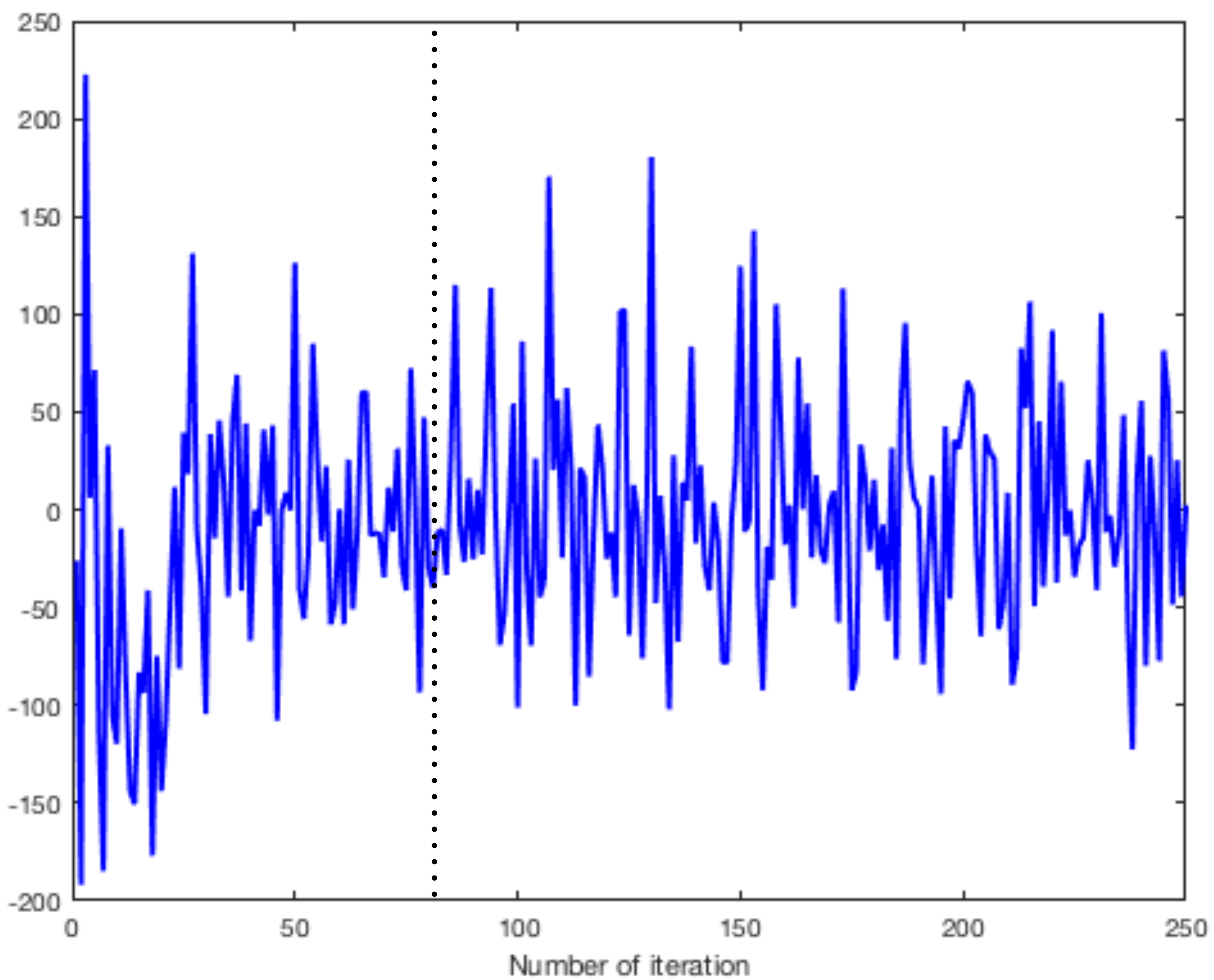


Weighted increment

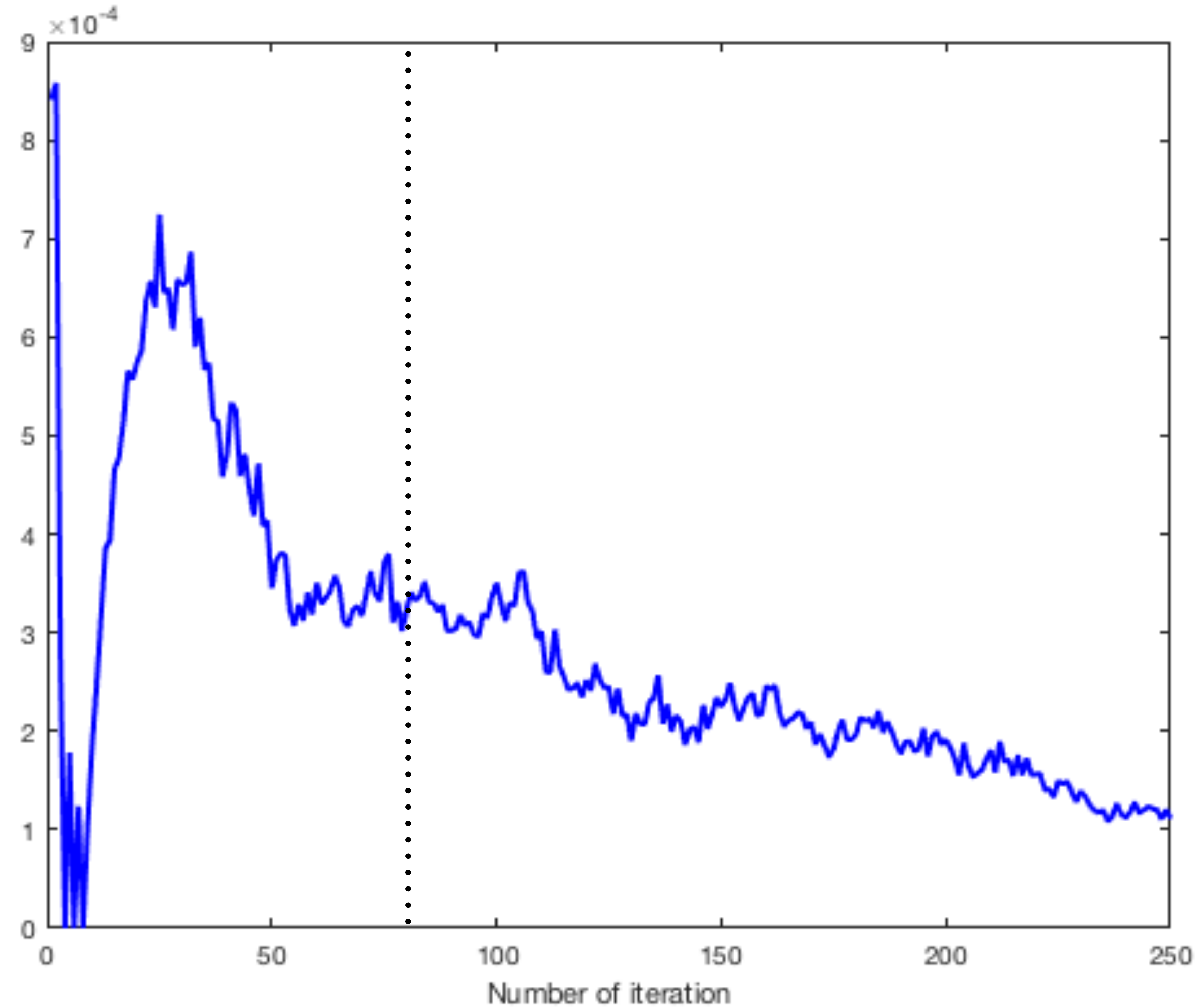


Intuition: Behaviour of the new weighted increment

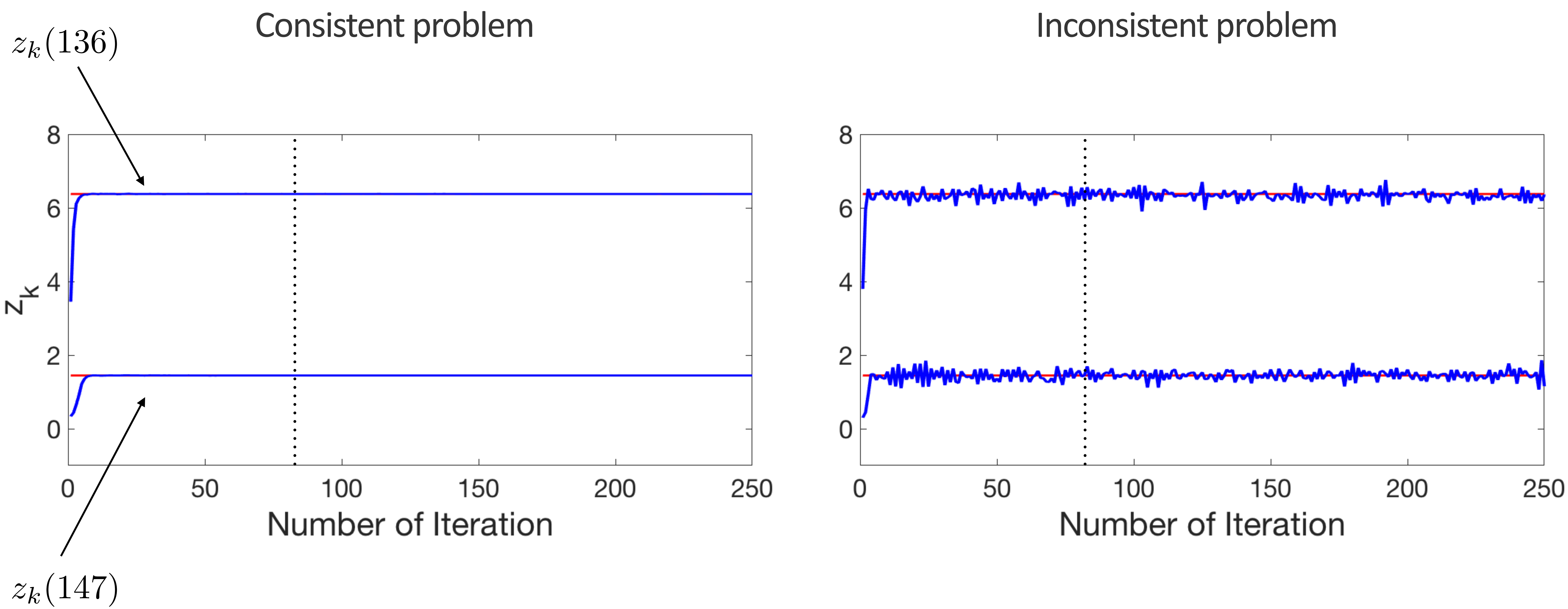
$$[A_k^T(A_k x_k - b_k)]_{147}$$



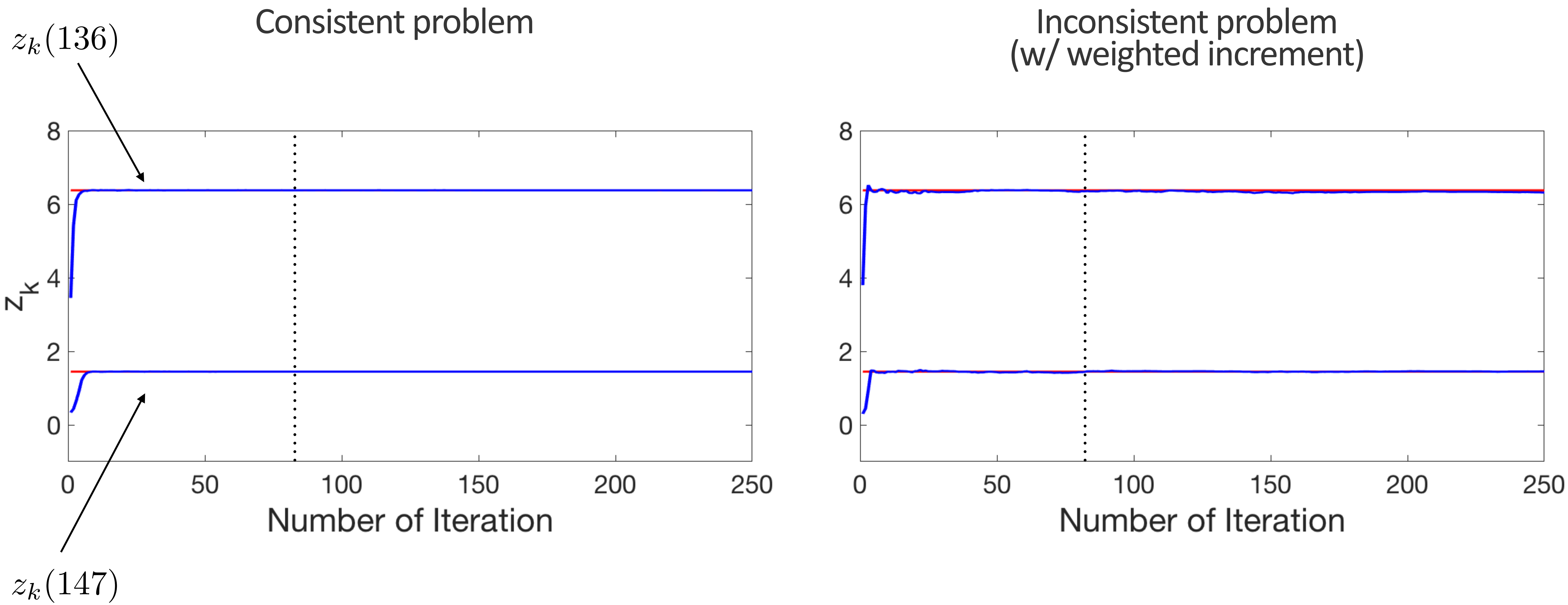
Weighted increment



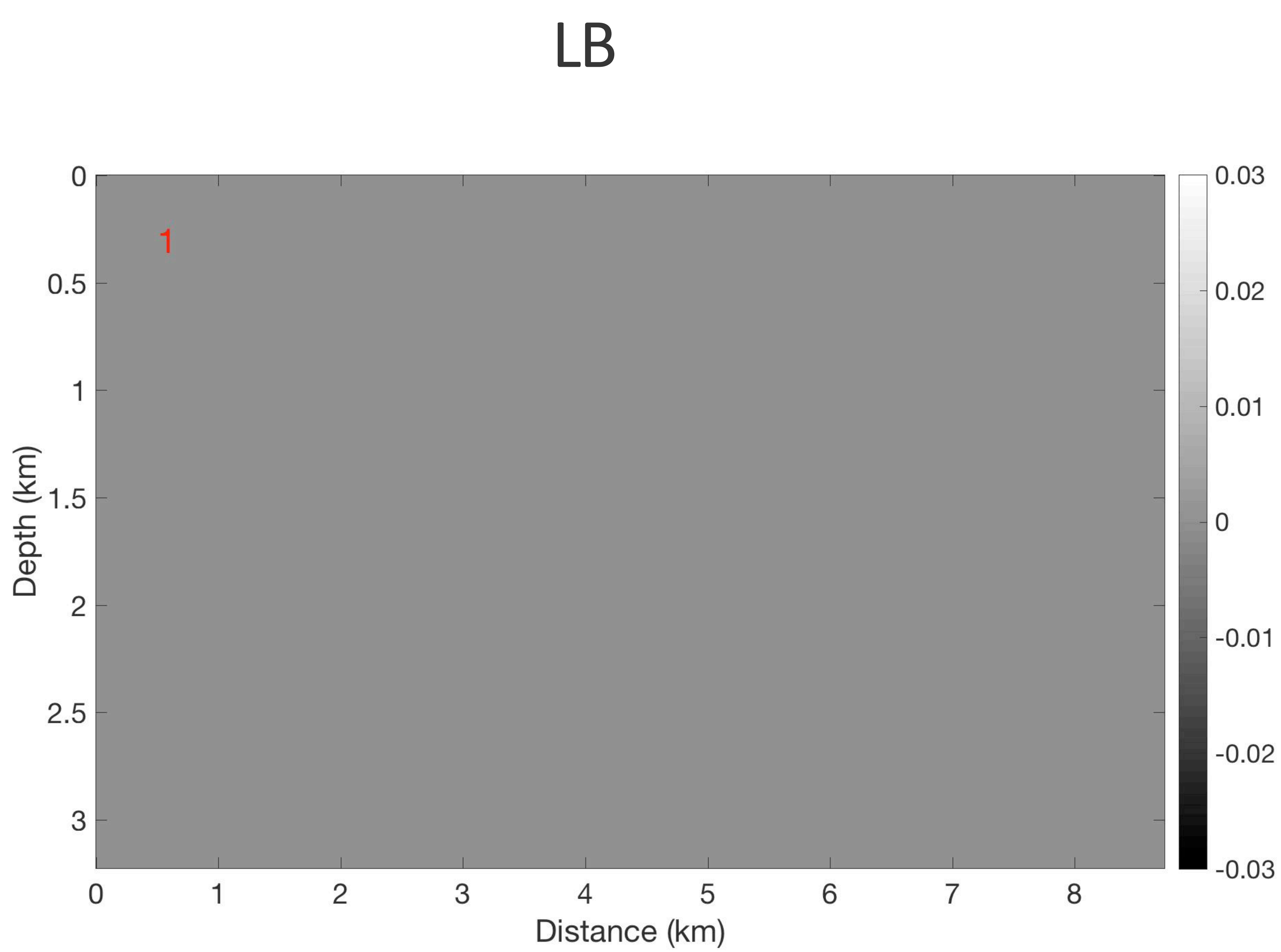
Intuition: Behaviour of the entries of the solution

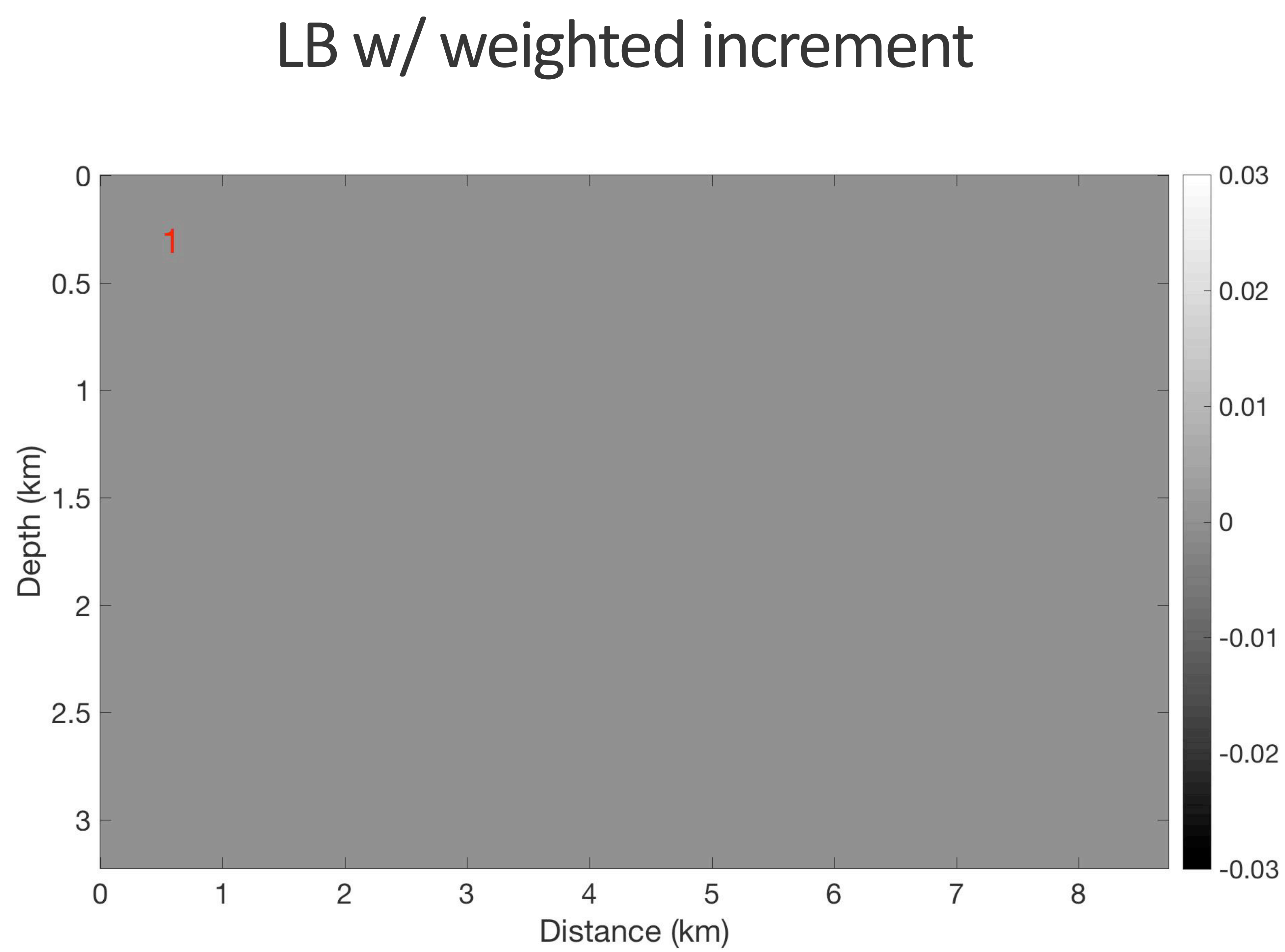


Intuition: Behaviour of the entries of the solution

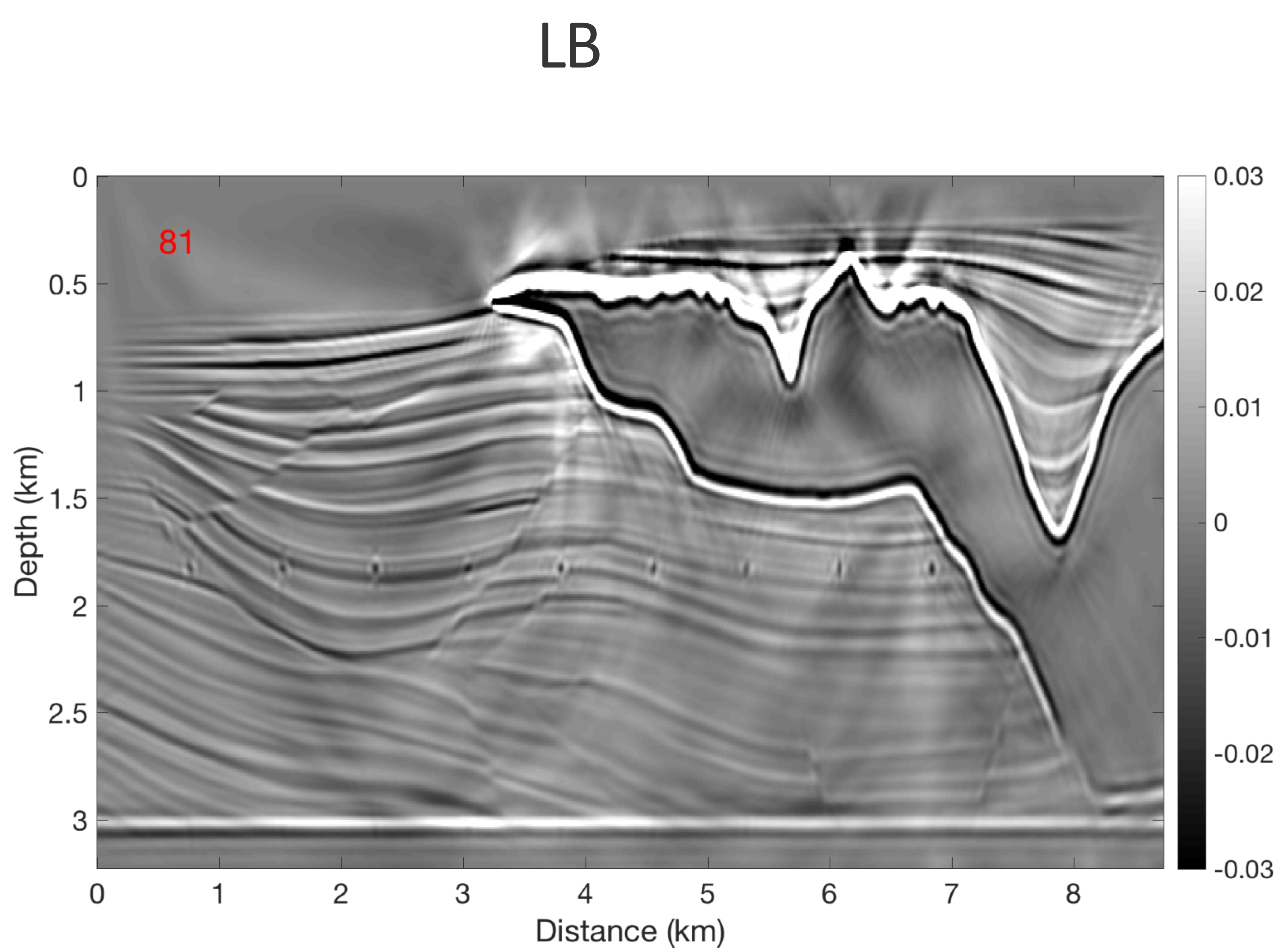


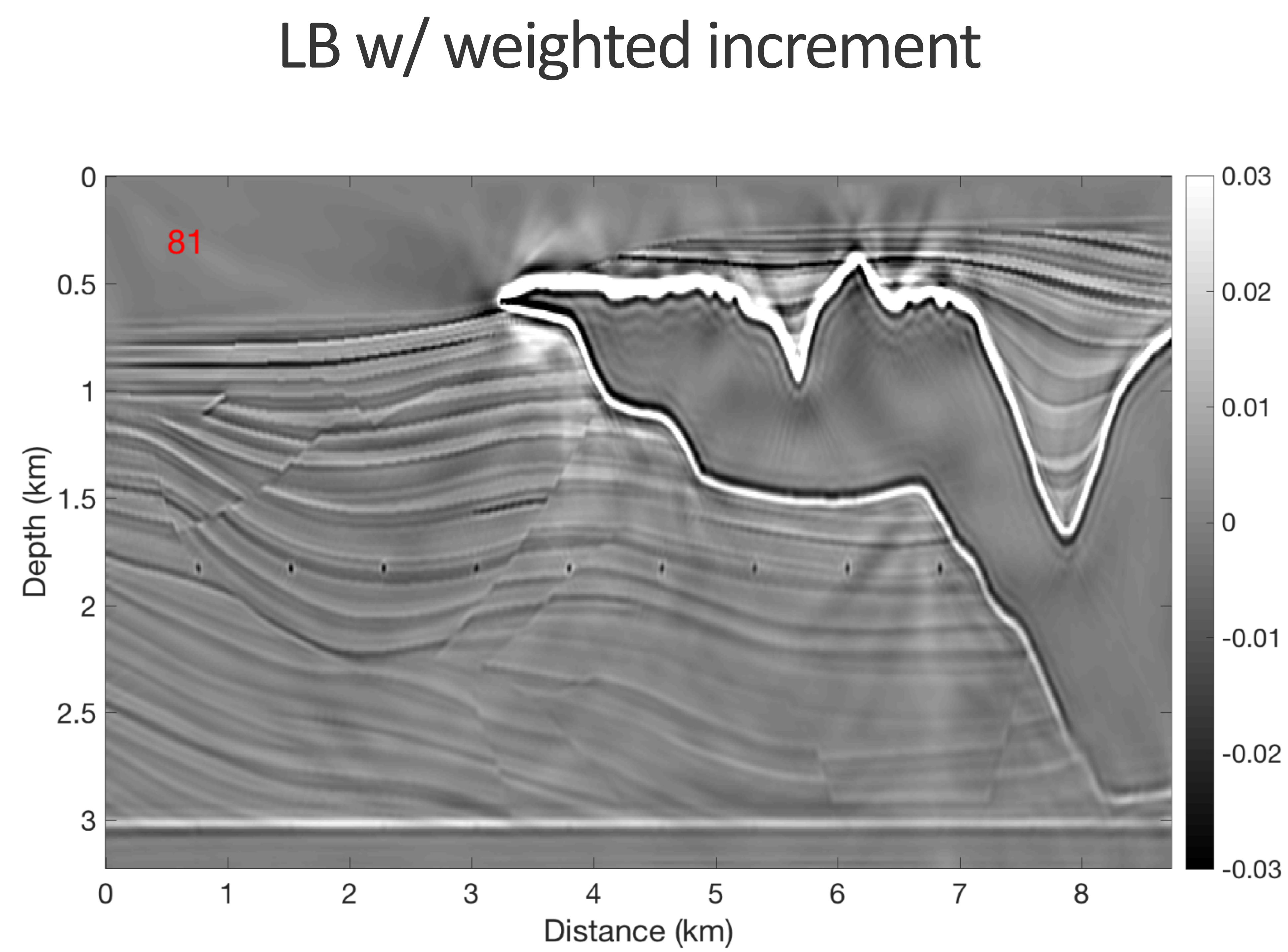
Effect on the LSRTM problem





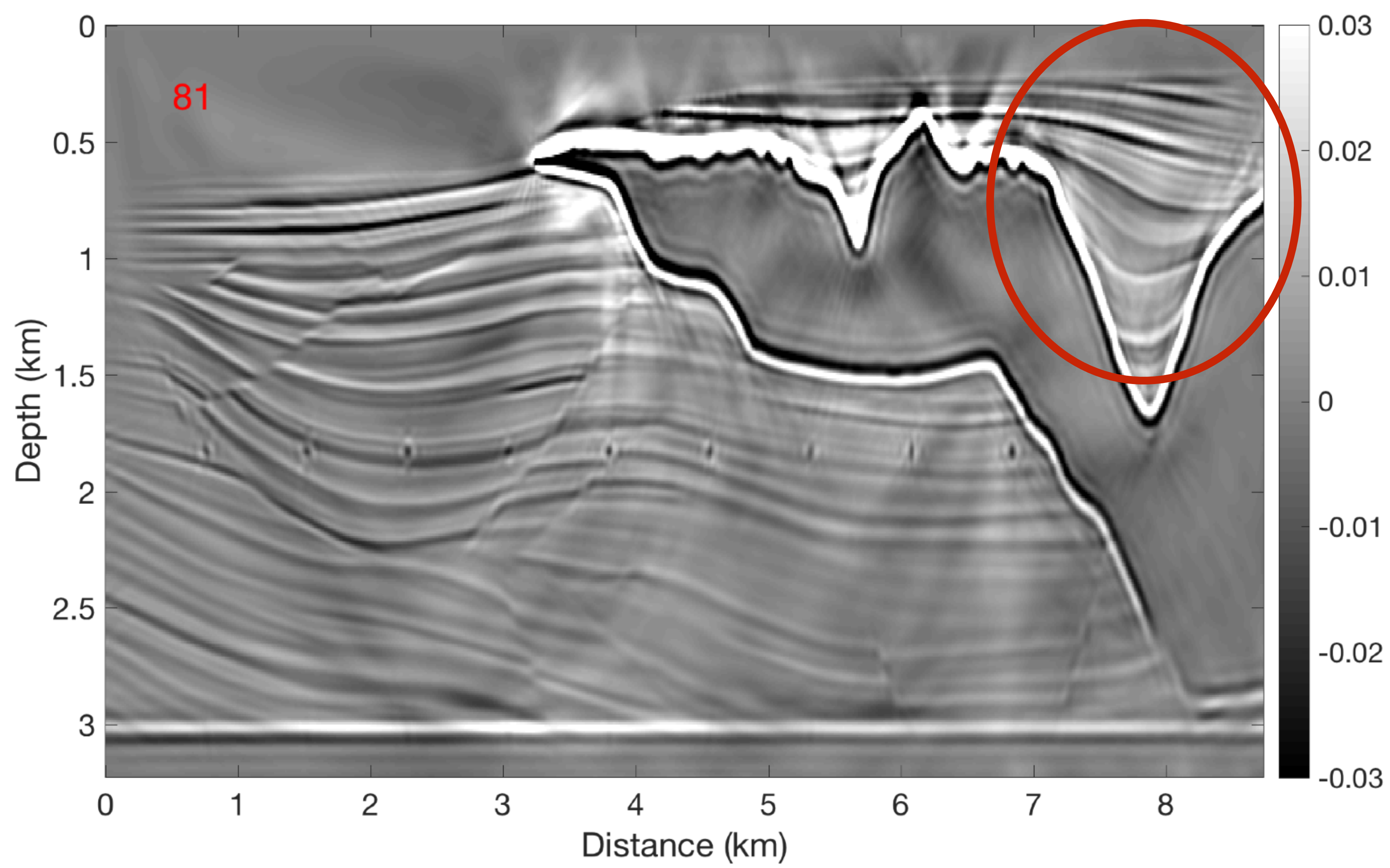
Effect on the LSRTM problem



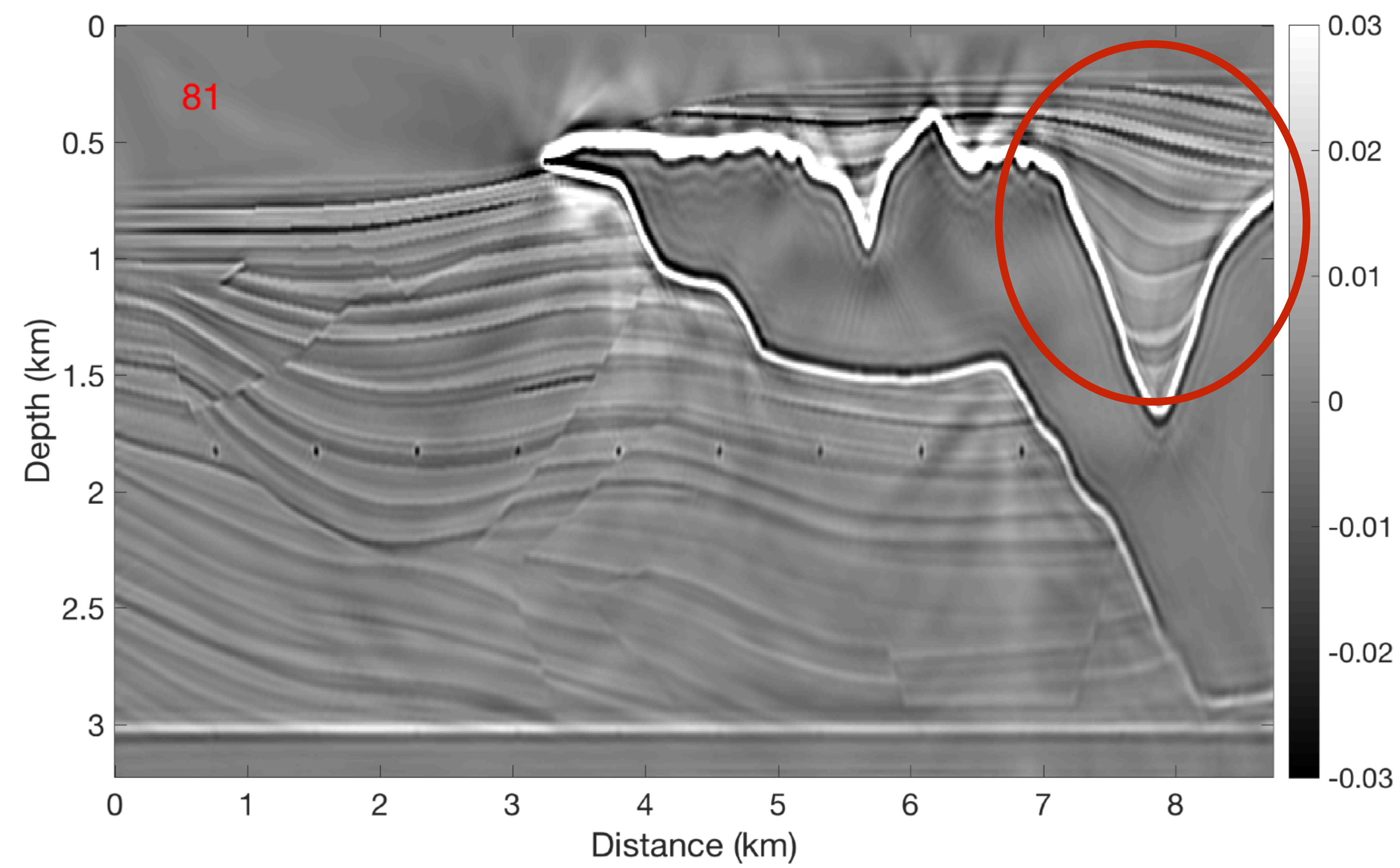


Effect on the LSRTM problem

LB

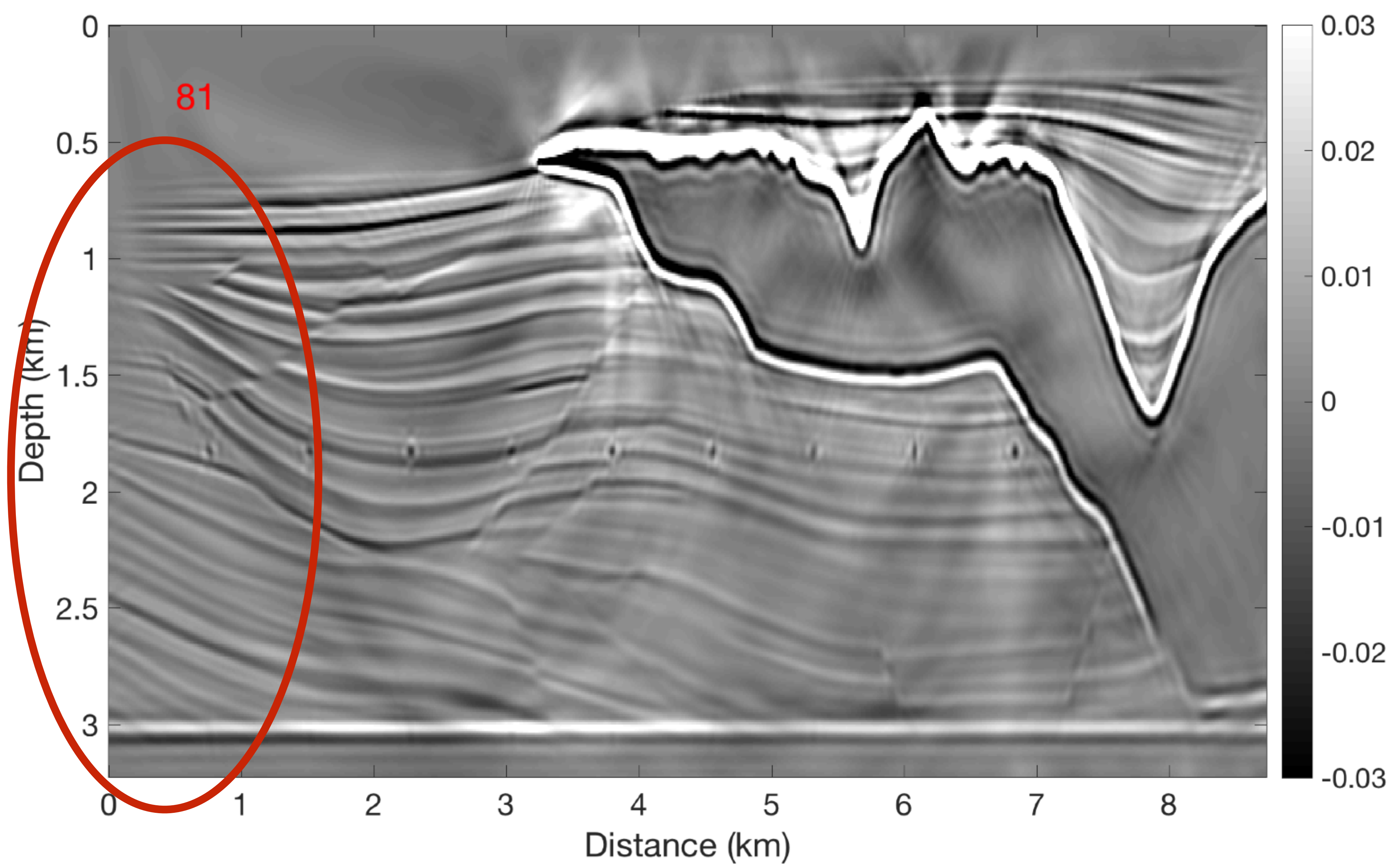


LB w/ weighted increment

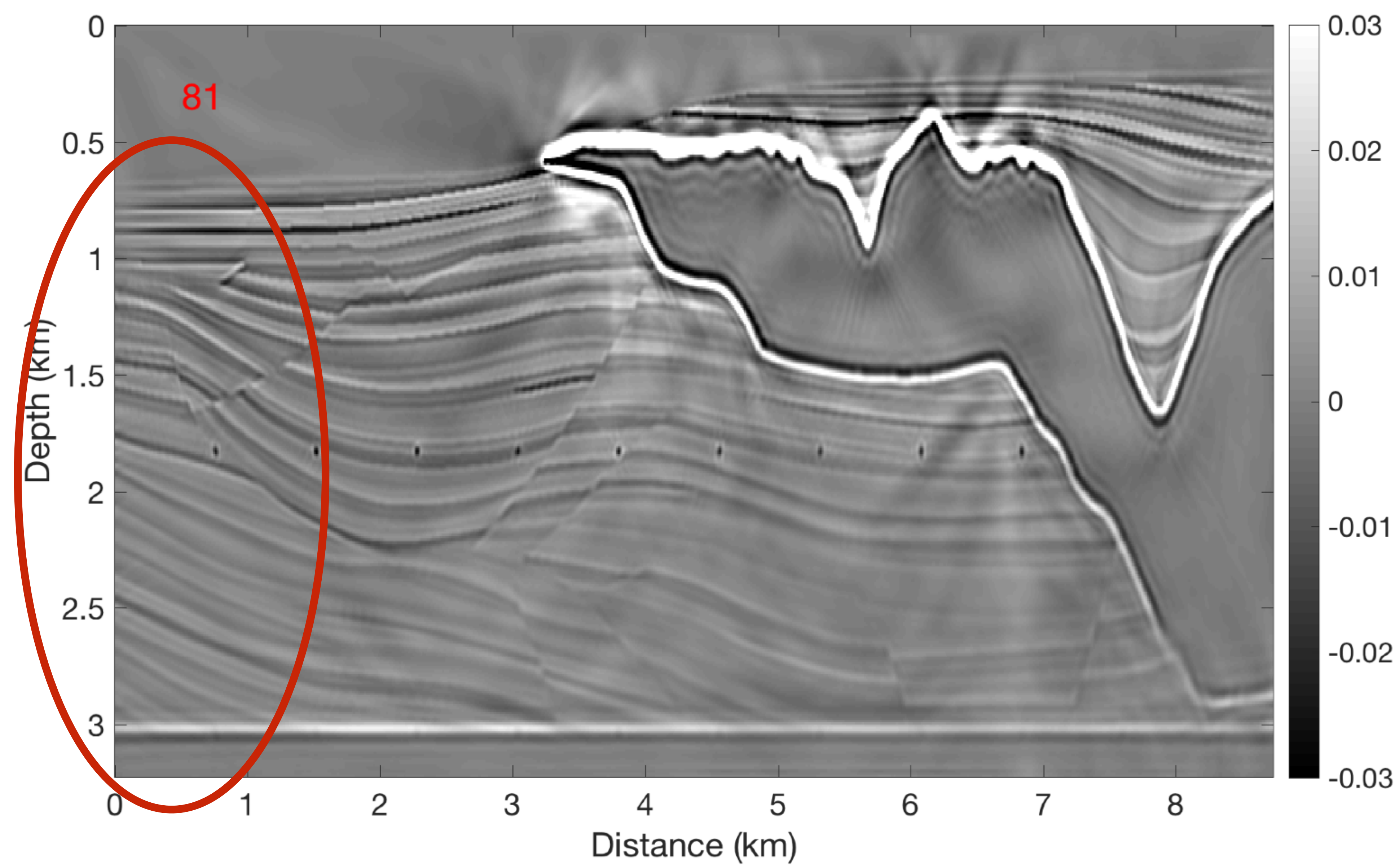


Effect on the LSRTM problem

LB

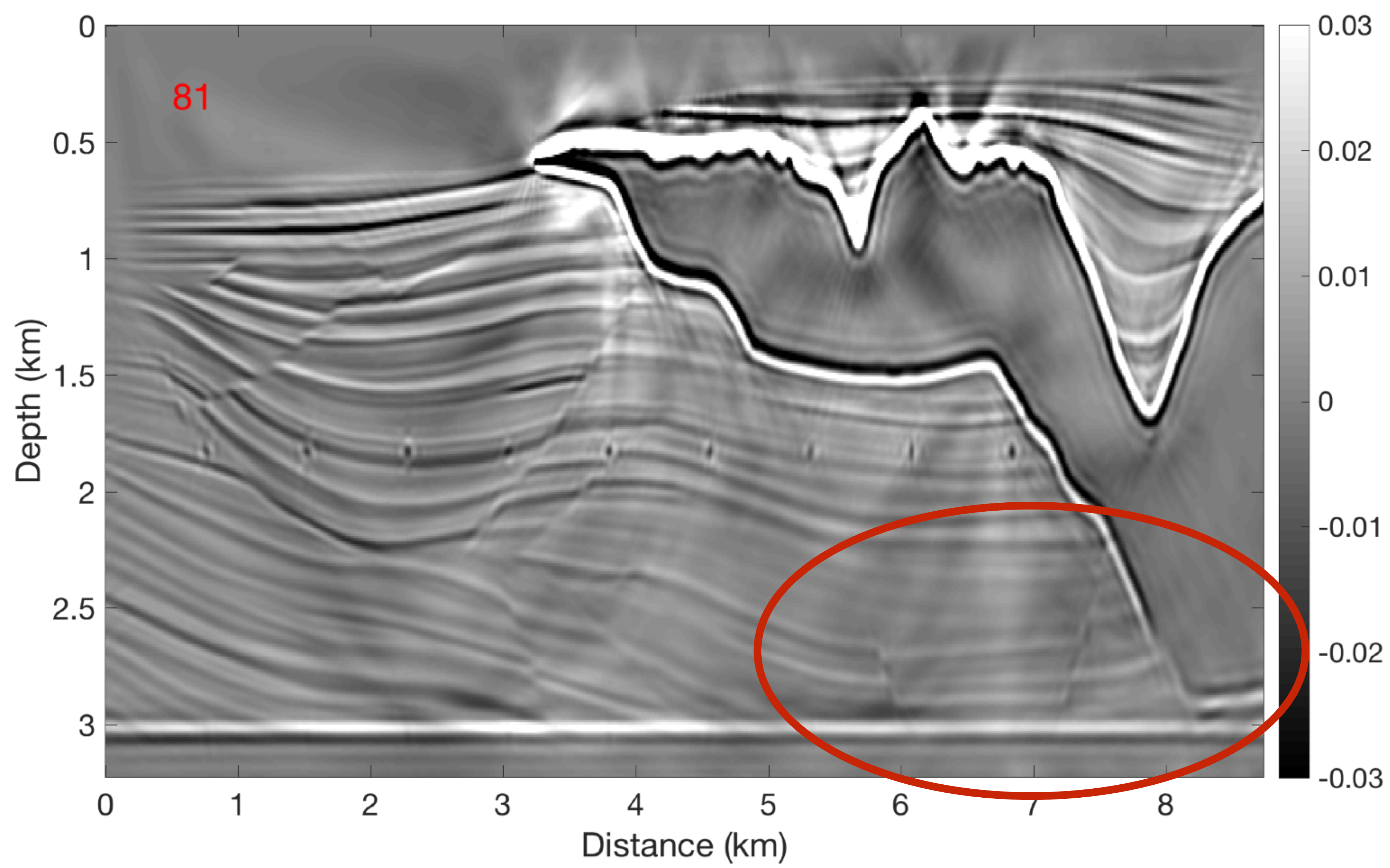


LB w/ weighted increment

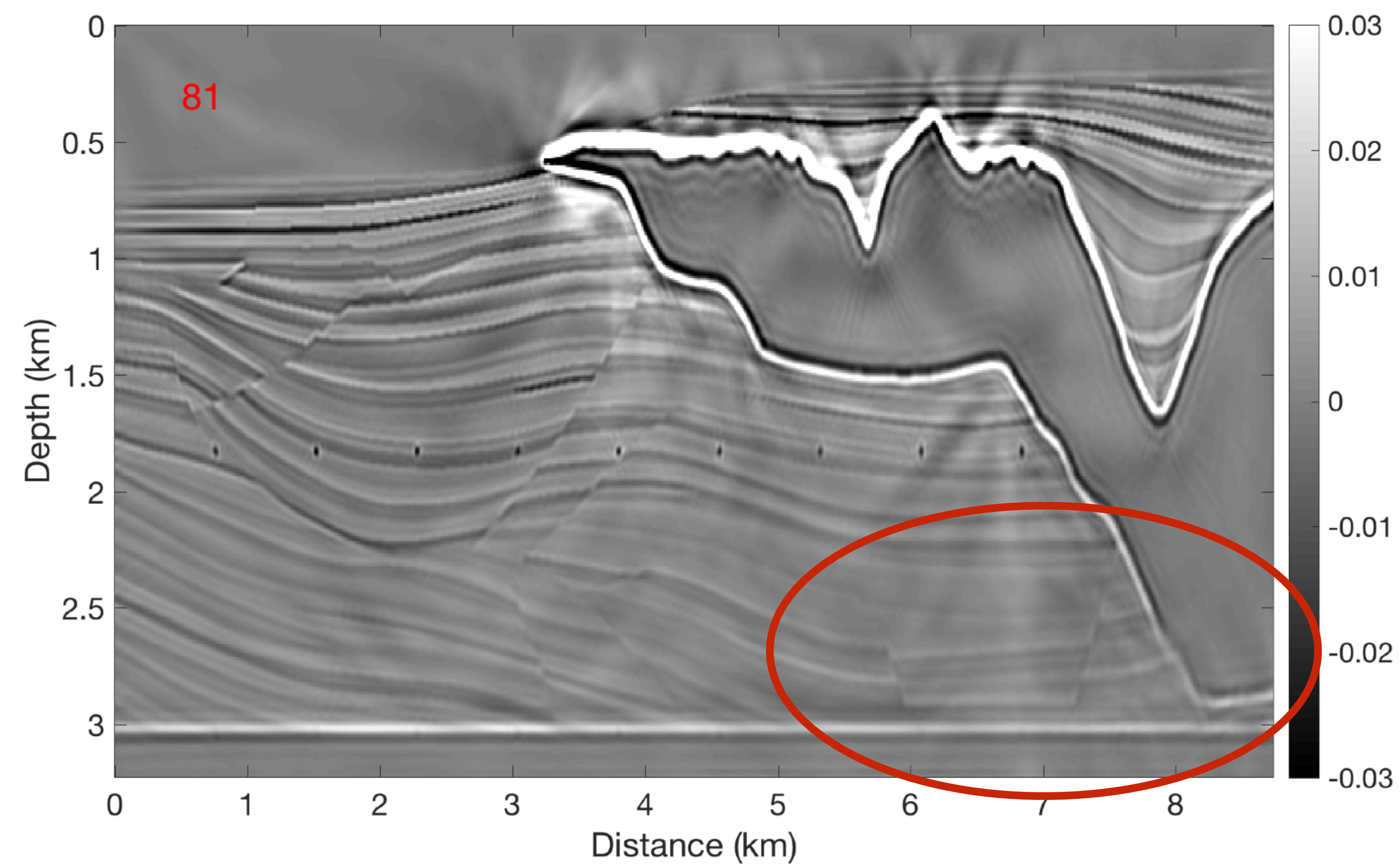


Effect on the LSRTM problem

LB

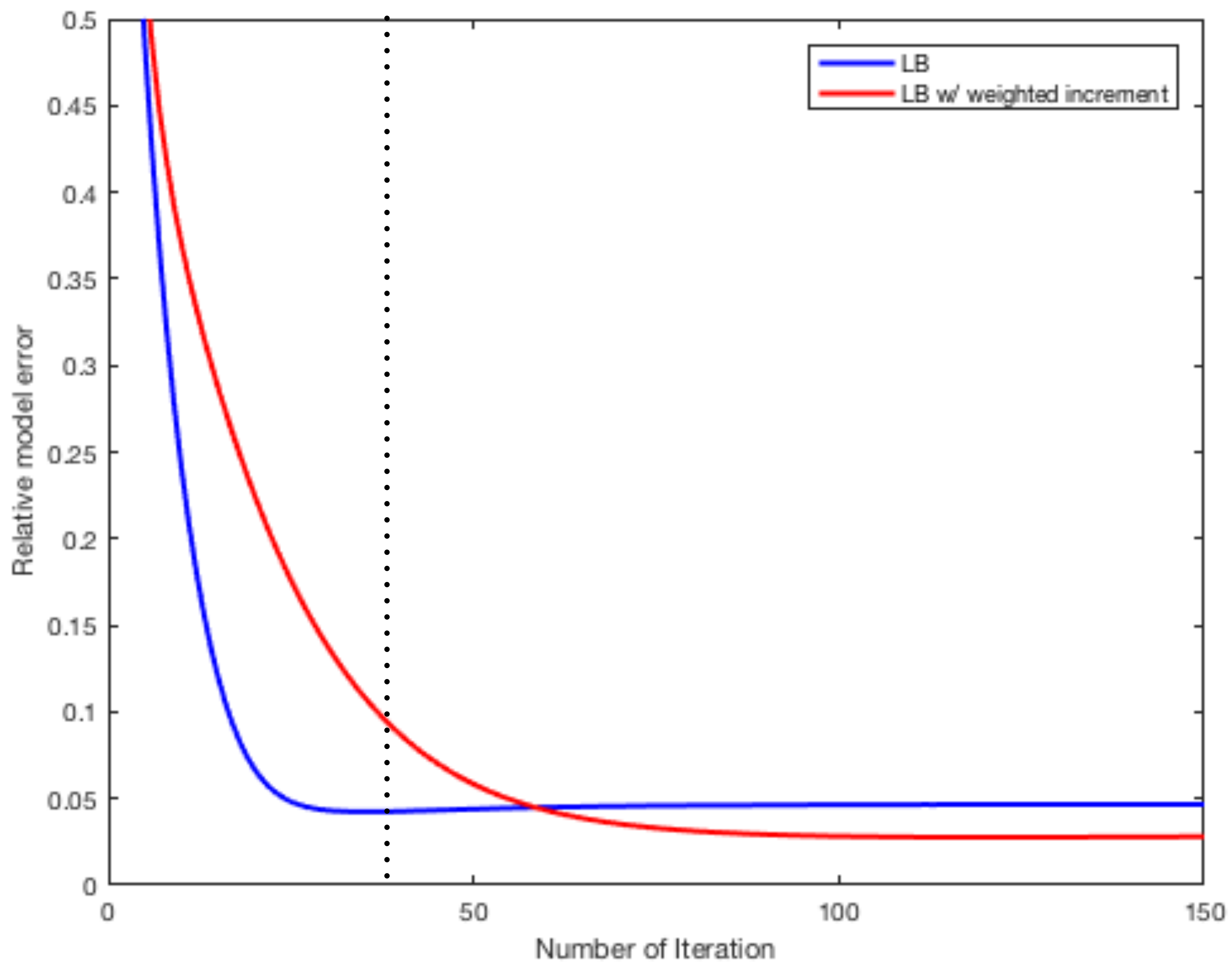


LB w/ weighted increment



Effect on large problems

Problem a



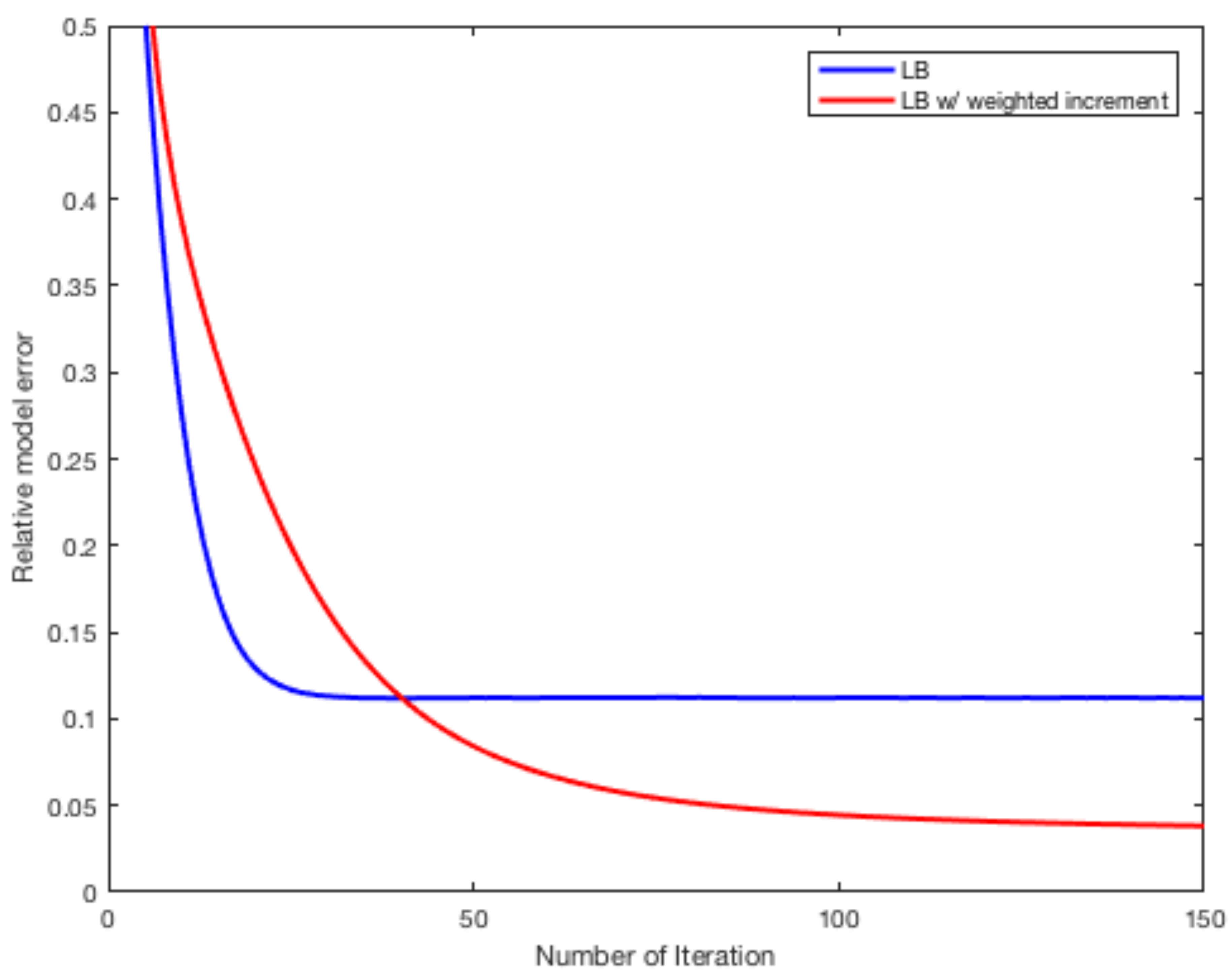
Problem a: $A \in \mathbb{R}^{10000000 \times 2040739}$

Problem b: $A \in \mathbb{R}^{70000000 \times 2040739}$

- The vector x corresponds to a known vector of curvelet coefficients
- $A_k \in \mathbb{R}^{500000 \times 2040739}$
- The signal to noise ratio for the data in both problems is the same.

Effect on large problems

Problem b



<u>Problem a:</u> $A \in \mathbb{R}^{10000000 \times 2040739}$

<u>Problem b:</u> $A \in \mathbb{R}^{70000000 \times 2040739}$

- The vector x corresponds to a known vector of curvelet coefficients
- $A_k \in \mathbb{R}^{500000 \times 2040739}$
- The signal to noise ratio for the data in both problems is the same.

Latest result

Remarks:

- The projection help us avoid overfitting the noise
- The weighted increment help us avoid cycling of the solution

Latest result

Remarks:

- The projection help us avoid overfitting the noise
- The weighted increment help us avoid cycling of the solution

Question: What if we combine projection and weighted increment?

SLIM

Latest result

LB method

$$
\begin{aligned}
z_{k+1} &= z_k - t_k A_k^T (A_k x_k - b_k) \\
x_{k+1} &= S_\lambda(z_{k+1}),
\end{aligned}
$$

where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\text{sign}(z_k)$$

$$t_k = \frac{||A_k x_k - b_k||_2^2}{||A_k^T (A_k x_k - b_k)||_2^2}$$

LB method w/ L2 ball projection and weighted increment

$$
\begin{aligned}
z_{k+1} &= z_k - \tau_k \odot A_k^T \Pi_\sigma (A_k x_k - b_k) \\
x_{k+1} &= S_\lambda(z_{k+1}),
\end{aligned}
$$

where

$$S_\lambda(z_k) = \max(|z_k| - \lambda, 0)\text{sign}(z_k)$$

and

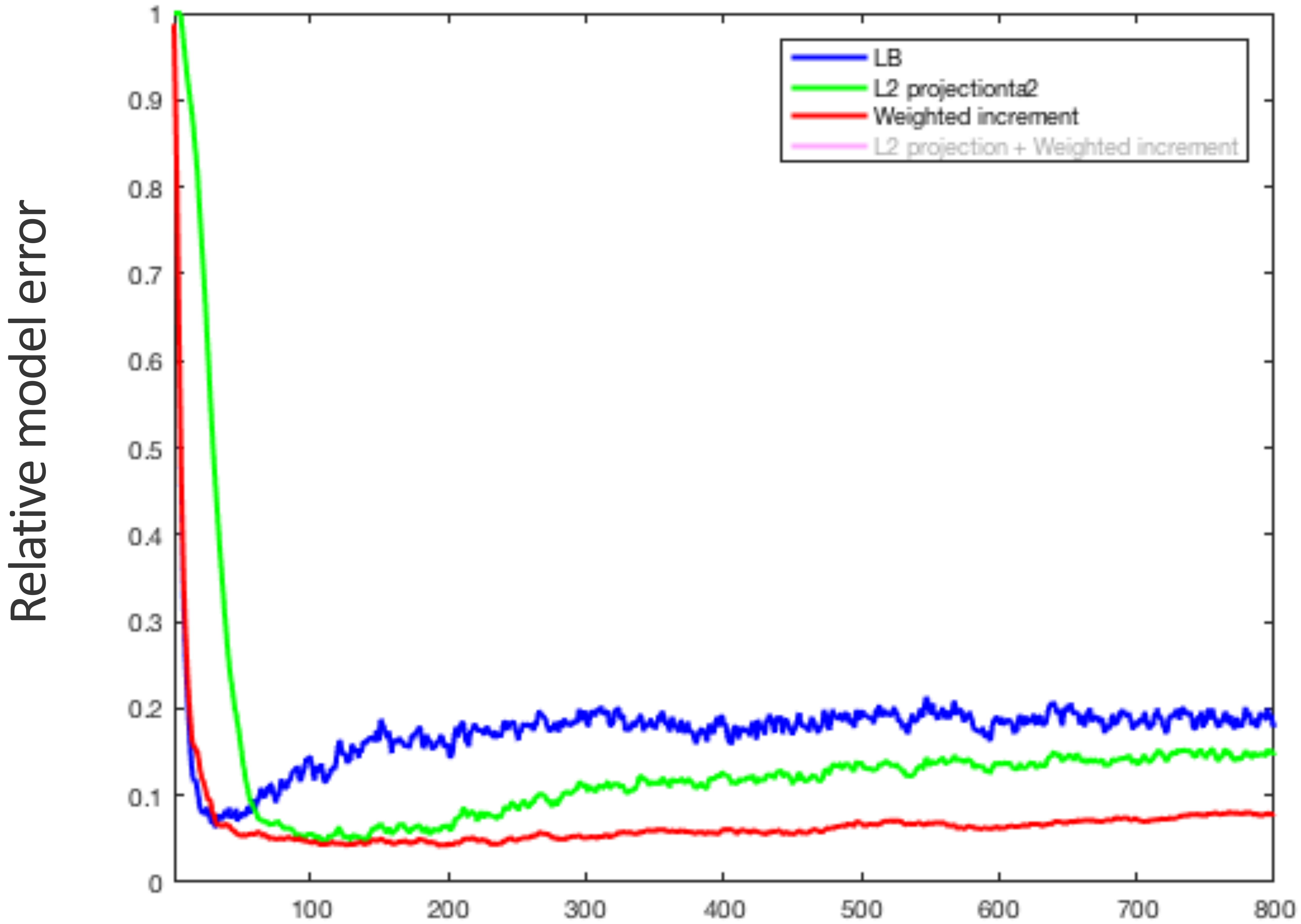
$$\tau_k^i = t_k \frac{|\sum_{j=1}^{k} \text{sign}([A_j^T (A_j x_j - b_j)]_i)|}{k}$$

Note:

$$\Pi_\sigma(x) = \max\{1 - \frac{\sigma}{||x||_2}\}(x)$$

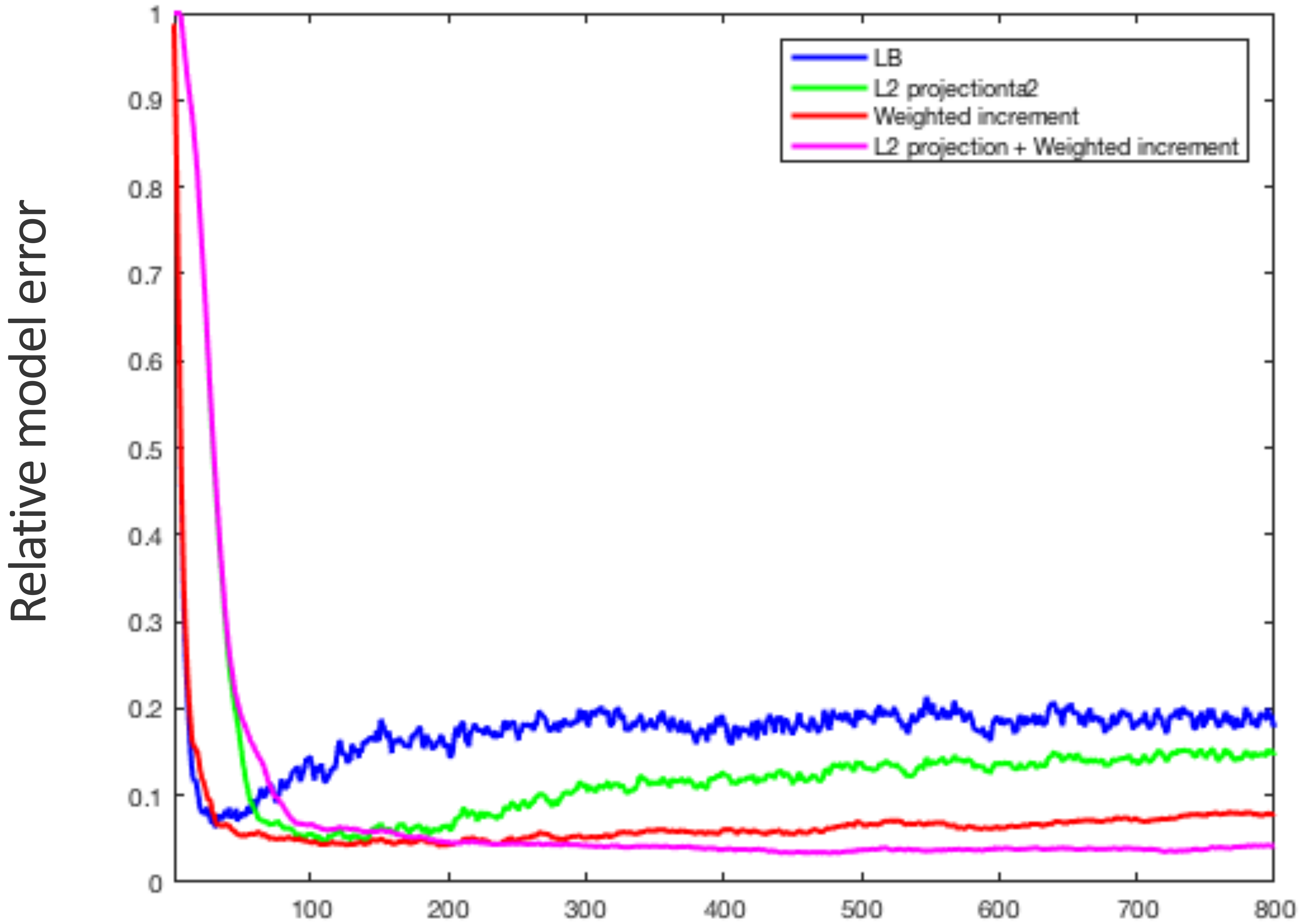
the projection on to l_2-norm ball

Latest result



10 data passes

Latest result



10 data passes

Acknowledgement

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.