# Resolution Analysis in Seismic Imaging Using the Kronecker-Factored Hessian

Gian Matharu, Wenlei Gao and Mauricio D. Sacchi

SIAM: Mathematical & Computational Issues in the Geosciences, Houston

University of Alberta | Signal Analysis & Imaging Group | NSERC CRSNG | Alberta Innovates

# Outline

- Motivation: Resolution and uncertainty analysis in FWI
- Kronecker-factored Hessian
- Numerical examples
  - I. Local resolution analysis
  - II. Linearized Bayesian inversion
- Conclusion

# Full waveform inversion

Regularized least-squares waveform inversion

$$\underset{\mathbf{m}}{\arg\min}\; J(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{n} \|\mathscr{P}\mathbf{u}_i(\mathbf{m}) - \mathbf{d}_i\|^2 + R(\mathbf{m})$$

Subject to the PDE constraint

$$\mathbf{L}(\mathbf{m})\mathbf{u}_i = \mathbf{s}_i$$

Wave equation operator

External source

# Full waveform inversion

Regularized least-squares waveform inversion

$$\underset{\mathbf{m}}{\arg\min}\ J(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{n} \|\mathscr{P}\mathbf{u}_i(\mathbf{m}) - \mathbf{d}_i\|^2 + R(\mathbf{m})$$

Updates performed via gradient-based optimization

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \nu^k \delta\mathbf{m}^k$$

Gradients computed using adjoint-state method (Plessix, 2006)

# Estimating uncertainties

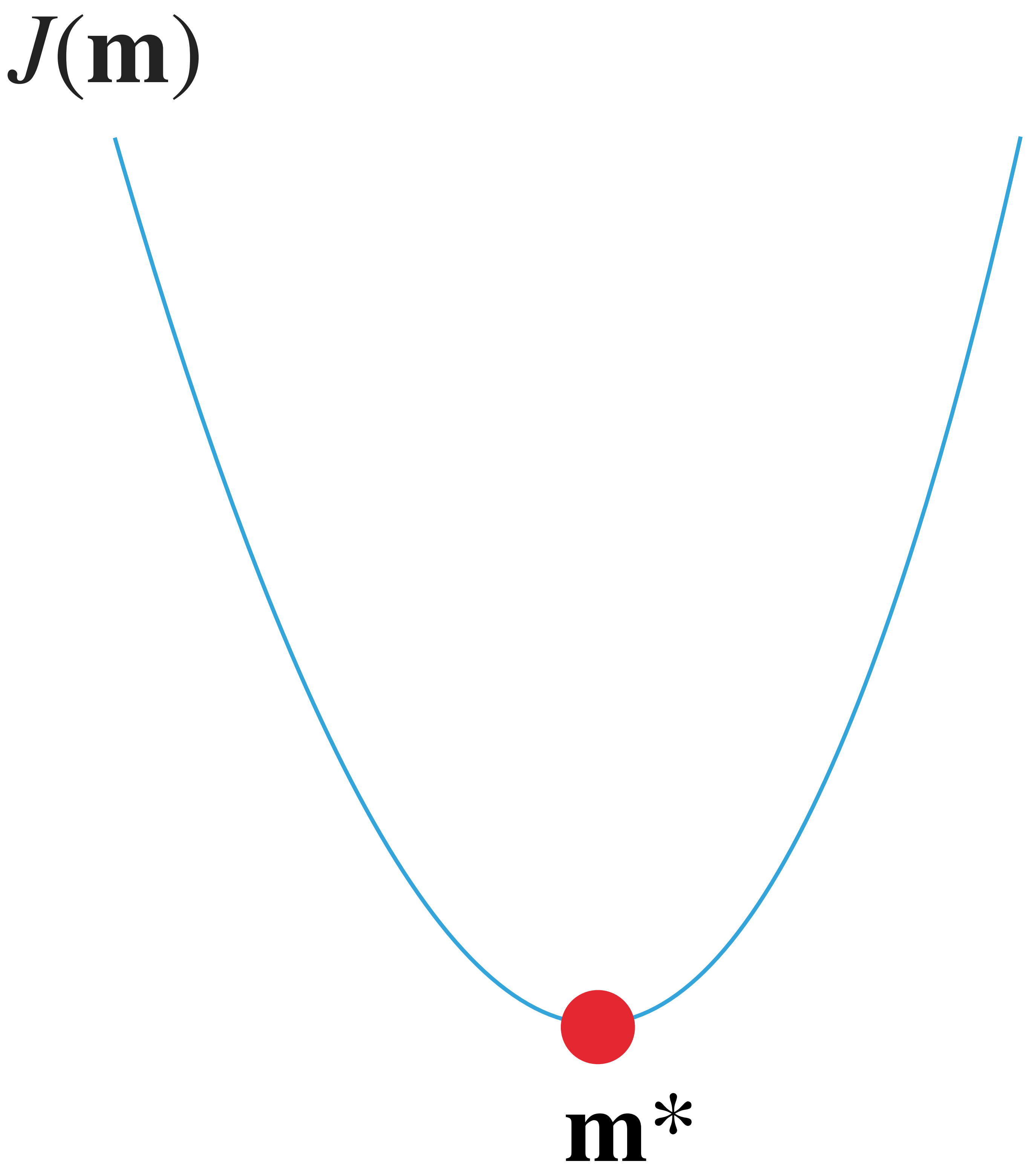
*Assuming convergence, how well constrained are the inverted models for a given dataset?*

*Focus on two interpretations of uncertainty analysis*

- Local resolution analysis
- Linearized Bayesian inversion

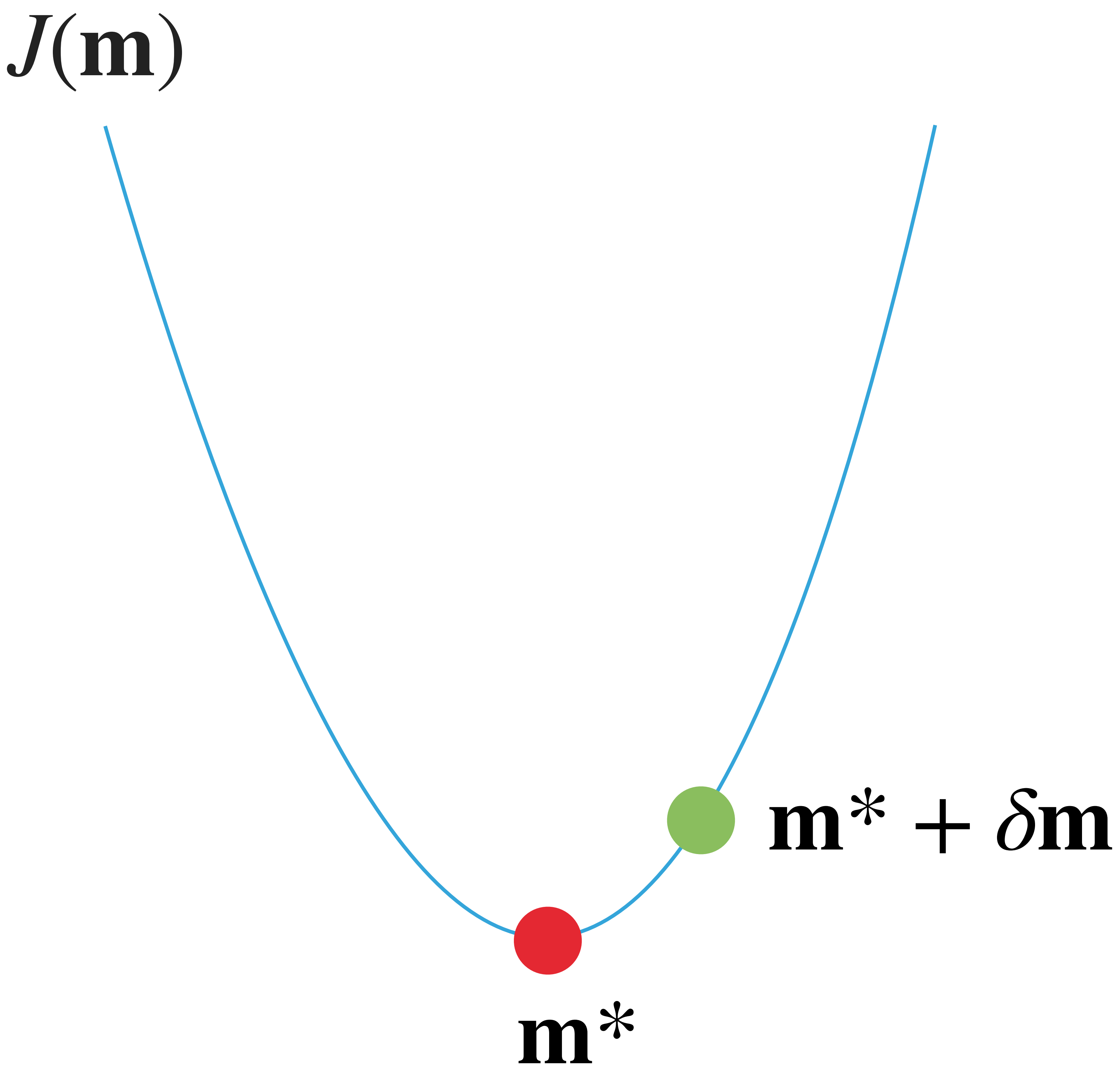
# I. Local resolution analysis

Follows the interpretation of Fichtner and Van Leeuwen (2015)



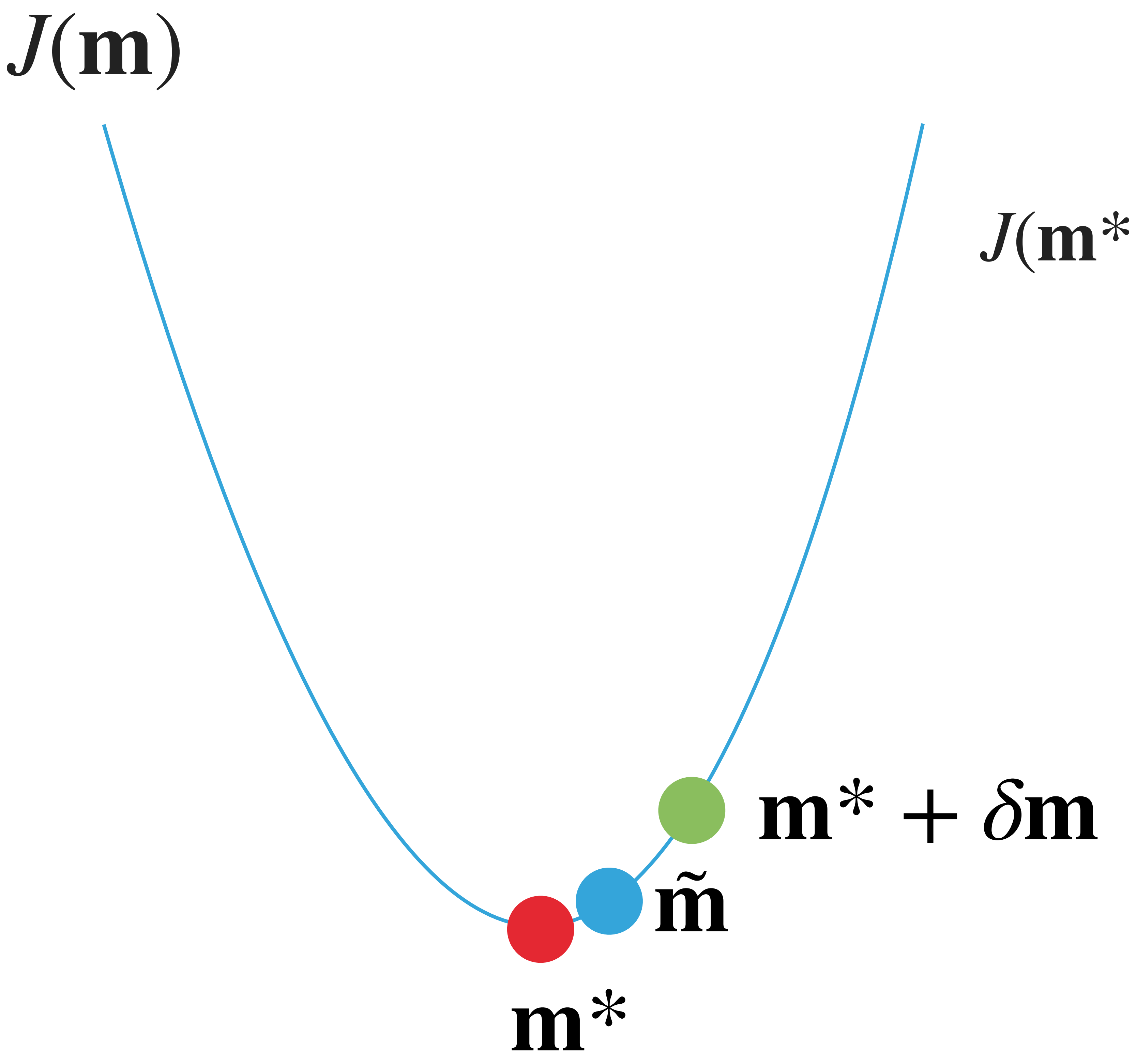
# I. Local resolution analysis

Follows the interpretation of Fichtner and Van Leeuwen (2015)



# I. Local resolution analysis

Follows the interpretation of Fichtner and Van Leeuwen (2015)

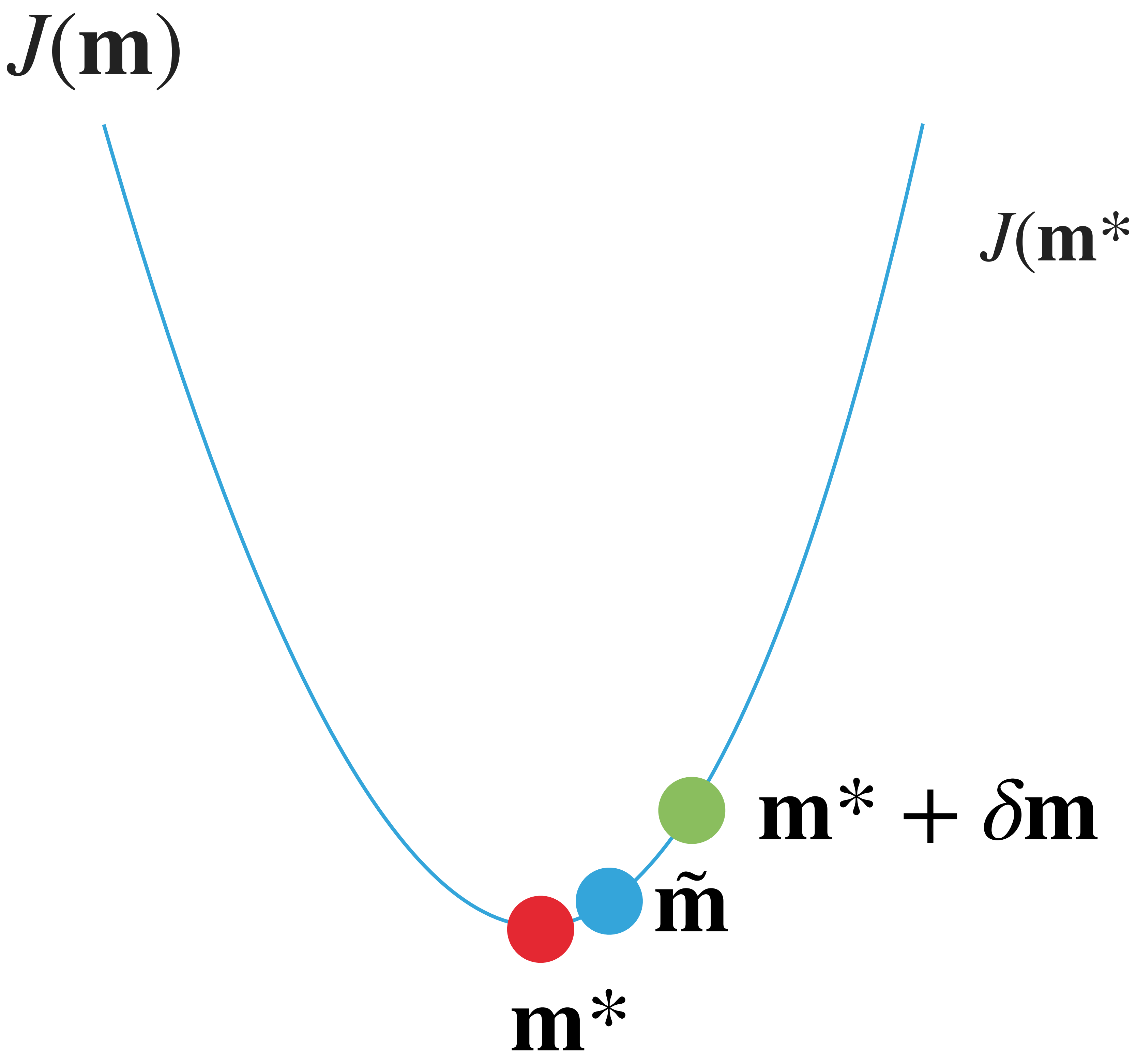



$$J(\mathbf{m}^* + \delta\mathbf{m}) = J(\mathbf{m}^*) + \cancel{\delta\mathbf{m}^T\mathbf{g}(\mathbf{m}^*)} + \frac{1}{2}\delta\mathbf{m}^T\mathbf{H}(\mathbf{m}^*)\delta\mathbf{m}$$

$$\tilde{\mathbf{m}} = (\mathbf{m}^* + \delta\mathbf{m}) - \nu\mathbf{H}\delta\mathbf{m}$$

# I. Local resolution analysis

Follows the interpretation of Fichtner and Van Leeuwen (2015)




$$J(\mathbf{m}^* + \delta\mathbf{m}) = J(\mathbf{m}^*) + \cancel{\delta\mathbf{m}^T\mathbf{g}(\mathbf{m}^*)} + \frac{1}{2}\delta\mathbf{m}^T\mathbf{H}(\mathbf{m}^*)\delta\mathbf{m}$$

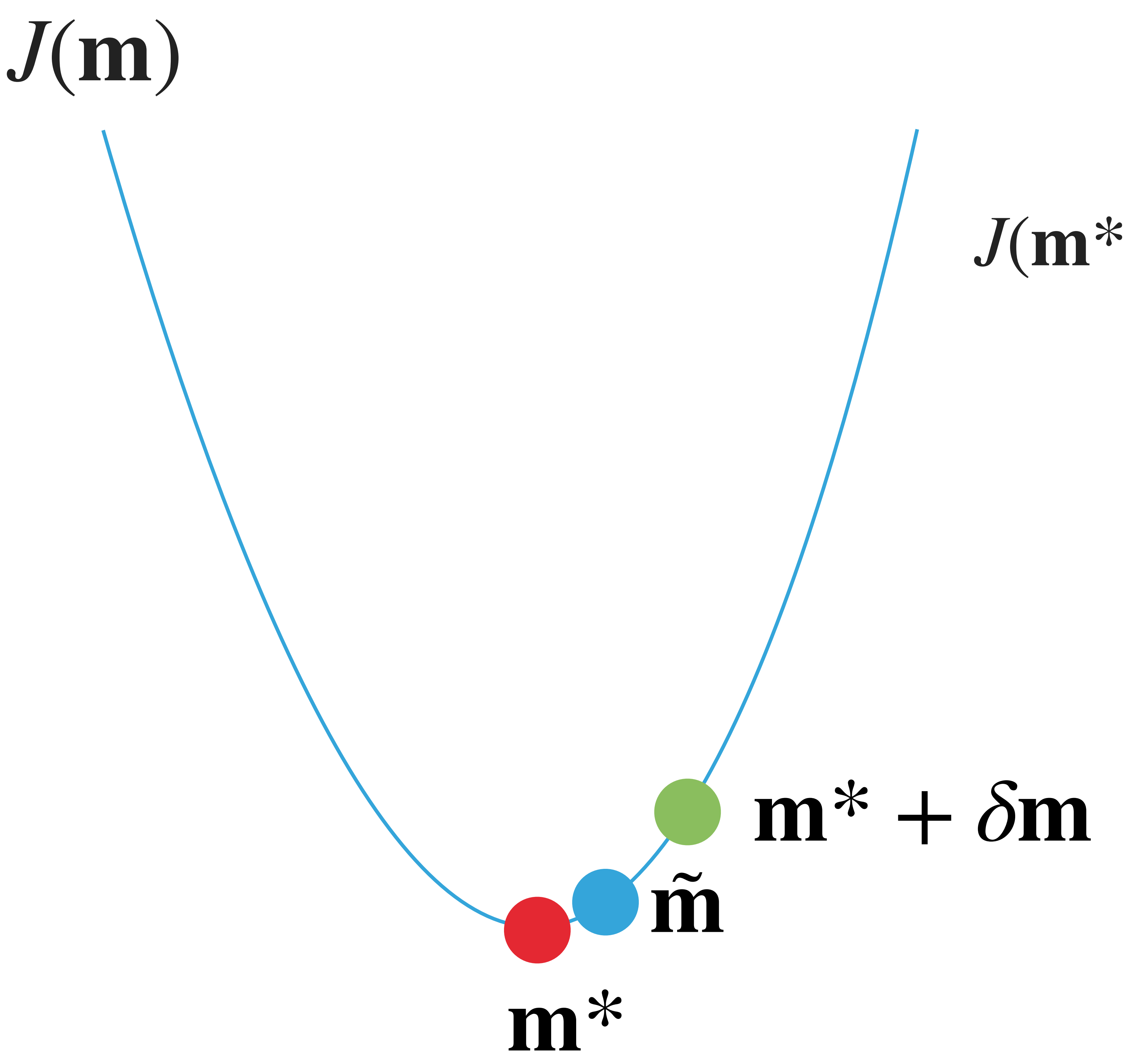
$$\tilde{\mathbf{m}} = (\mathbf{m}^* + \delta\mathbf{m}) - \nu\mathbf{H}\delta\mathbf{m}$$

$$\Downarrow$$

True perturbation smeared by Hessian

# I. Local resolution analysis

Follows the interpretation of Fichtner and Van Leeuwen (2015)




$$J(\mathbf{m}^* + \delta\mathbf{m}) = J(\mathbf{m}^*) + \cancel{\delta\mathbf{m}^T\mathbf{g}(\mathbf{m}^*)} + \frac{1}{2}\delta\mathbf{m}^T\mathbf{H}(\mathbf{m}^*)\delta\mathbf{m}$$

$$\tilde{\mathbf{m}} = (\mathbf{m}^* + \delta\mathbf{m}) - \nu\mathbf{H}\delta\mathbf{m}$$

True perturbation smeared by Hessian

Relies on second-order adjoint method

# II. Bayesian formulation

The Bayesian framework formulates the inverse problem as an inference problem (Tarantola, 2005)

$$\rho(\mathbf{m} \,|\, \mathbf{d}) \propto \rho(\mathbf{d} \,|\, \mathbf{m})\rho(\mathbf{m})$$

Posterior Likelihood Model prior

# II. Bayesian formulation

Assuming Gaussian priors,

$$\rho(\mathbf{m}\,|\,\mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{u}(\mathbf{m}) - \mathbf{d})^T C_{\mathbf{D}}^{-1}(\mathbf{u}(\mathbf{m}) - \mathbf{d}) - \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T C_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}_0)]$$

Optimization seeks to maximize log-likelihood

$$\arg\max_{\mathbf{m}} \ \log \rho(\mathbf{m}\,|\,\mathbf{d})$$

# II. Bayesian formulation

Assuming Gaussian priors,

$$\rho(\mathbf{m}\,|\,\mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{u}(\mathbf{m}) - \mathbf{d})^T C_{\mathbf{D}}^{-1}(\mathbf{u}(\mathbf{m}) - \mathbf{d}) - \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T C_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}_0)]$$

Equivalent to regularized least-squares

$$\arg\min_{\mathbf{m}} \ \frac{1}{2}\|(\mathbf{u}(\mathbf{m}) - \mathbf{d})\|^2_{C_{\mathbf{D}}^{-1}} + \frac{1}{2}\|(\mathbf{m} - \mathbf{m}_0)\|^2_{C_{\mathbf{m}}^{-1}}$$

# Posterior distribution

$$\mathbf{d} \approx \mathbf{u}(\mathbf{m}^*) + \mathbf{G}(\mathbf{m} - \mathbf{m}^*) \qquad (\mathbf{G} = \frac{\partial \mathbf{u}}{\partial \mathbf{m}})$$

Linearizing the modelling operator about **m*** results in a Gaussian posterior

$$\rho(\mathbf{m}\,|\,\mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{m} - \mathbf{m}^*)^T \tilde{C}_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}^*)]$$

With a posterior covariance defined as

$$\tilde{C}_{\mathbf{m}} = (\mathbf{H} + C_{\mathbf{m}}^{-1})^{-1}$$

# Sampling the posterior

Random samples from $\mathcal{N}(\mathbf{m}^*, \tilde{C}_{\mathbf{m}})$ can be drawn via

$$\mathbf{m}_s = \mathbf{m}^* + \tilde{C}_{\mathbf{m}}^{\frac{1}{2}}\mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0,1)$$

Cannot readily factor previous expression for $\tilde{C}_{\mathbf{m}}$

# Sampling the posterior

Random samples from $\mathcal{N}(\mathbf{m}^*, \tilde{C}_{\mathbf{m}})$ can be drawn via

$$\mathbf{m}_s = \mathbf{m}^* + \tilde{C}_{\mathbf{m}}^{\frac{1}{2}}\mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0,1)$$

Use the approach of Bui-Thanh et al. (2014)

$$\tilde{C}_{\mathbf{m}} = C_{\mathbf{m}}^{\frac{1}{2}}(C_{\mathbf{m}}^{\frac{1}{2}}\mathbf{H}C_{\mathbf{m}}^{\frac{1}{2}} + I)^{-1}C_{\mathbf{m}}^{\frac{1}{2}}$$

# Sampling the posterior

Random samples from $\mathcal{N}(\mathbf{m}^*, \tilde{C}_{\mathbf{m}})$ can be drawn via

$$\mathbf{m}_s = \mathbf{m}^* + \tilde{C}_{\mathbf{m}}^{\frac{1}{2}}\mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0,1)$$

Use the approach of Bui-Thanh et al. (2014)

$$\tilde{C}_{\mathbf{m}} = C_{\mathbf{m}}^{\frac{1}{2}}(\boxed{C_{\mathbf{m}}^{\frac{1}{2}}\mathbf{H}C_{\mathbf{m}}^{\frac{1}{2}}} + I)^{-1}C_{\mathbf{m}}^{\frac{1}{2}}$$

Approximate the following with a low-rank approximation

# Sampling the posterior

Random samples from $\mathcal{N}(\mathbf{m}^*, \tilde{C}_\mathbf{m})$ can be drawn via

$$\mathbf{m}_s = \mathbf{m}^* + \tilde{C}_\mathbf{m}^{\frac{1}{2}}\mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0,1)$$

Use the approach of Bui-Thanh et al. (2014)

$$\tilde{C}_\mathbf{m} = C_\mathbf{m}^{\frac{1}{2}}(\boxed{C_\mathbf{m}^{\frac{1}{2}}\mathbf{H}C_\mathbf{m}^{\frac{1}{2}}} + I)^{-1}C_\mathbf{m}^{\frac{1}{2}}$$

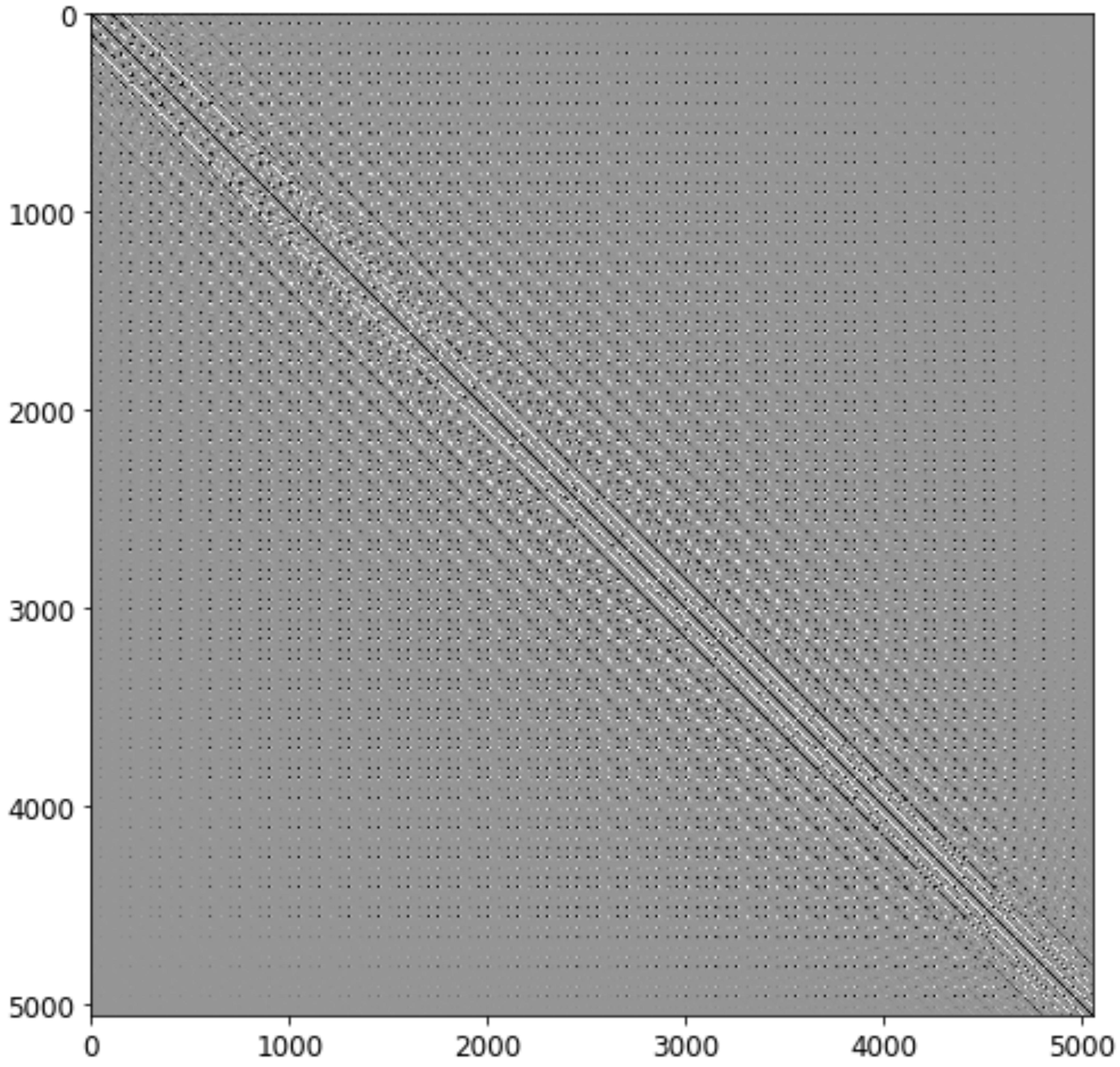
Lanczos iterations require Hessian-vector products

# Fast Hessian-vector products

- Outline two approaches to resolution/uncertainty analysis that feature the Hessian
- Hessian-vector products require costly PDE solves
- Propose a factorization of the Hessian in terms of a superposition of Kronecker products.
- Hessian-vector products involve operations on small matrices

# Kronecker-Factored Hessian

# Structure of the Hessian



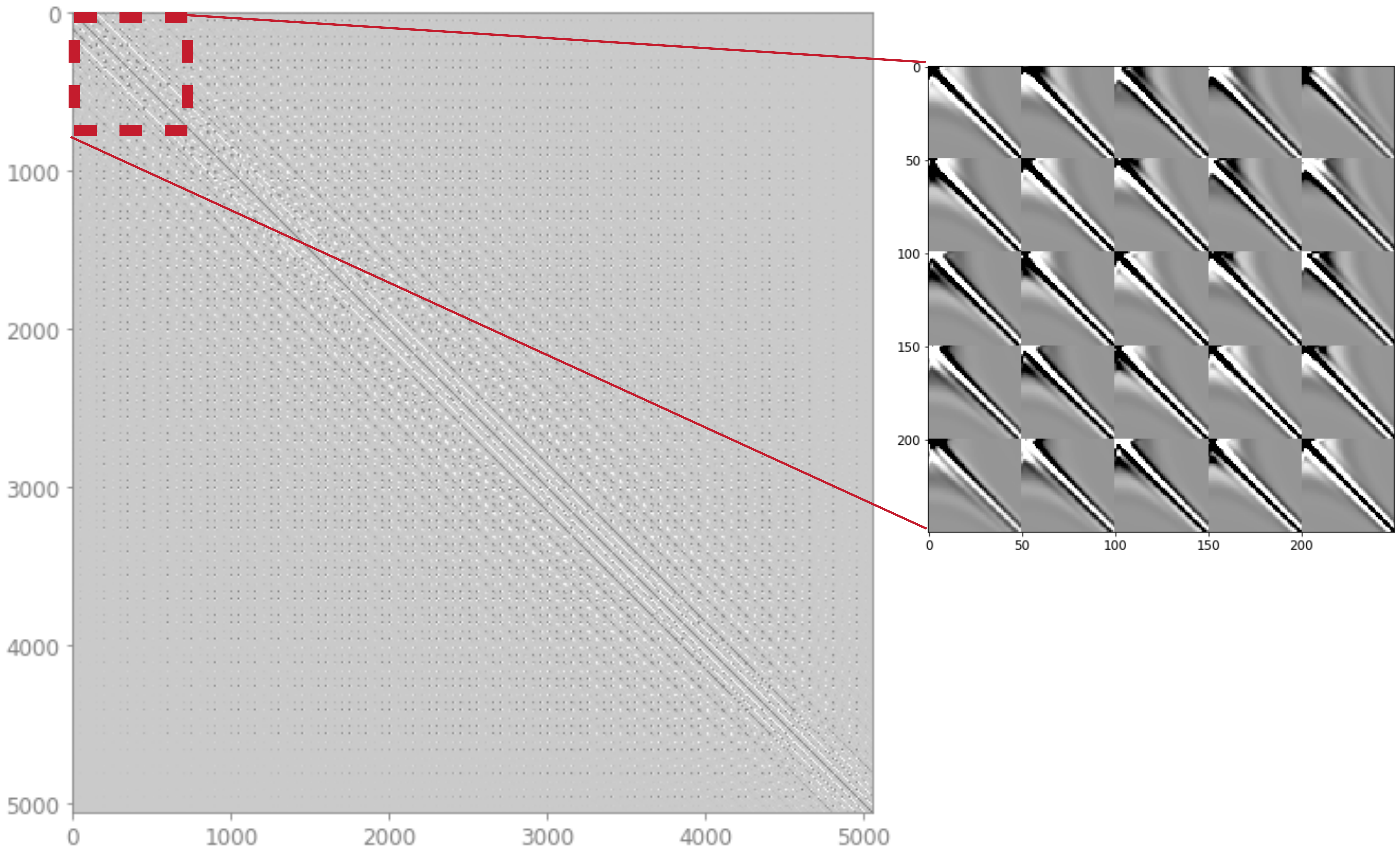

Hessian has a block-banded diagonal structure

Elements typically decay away from the diagonal

**Dimensions:** $n_x n_z$ x $n_x n_z$

(Computed in a homogeneous model of size 101 x 50)

# Structure of the Hessian

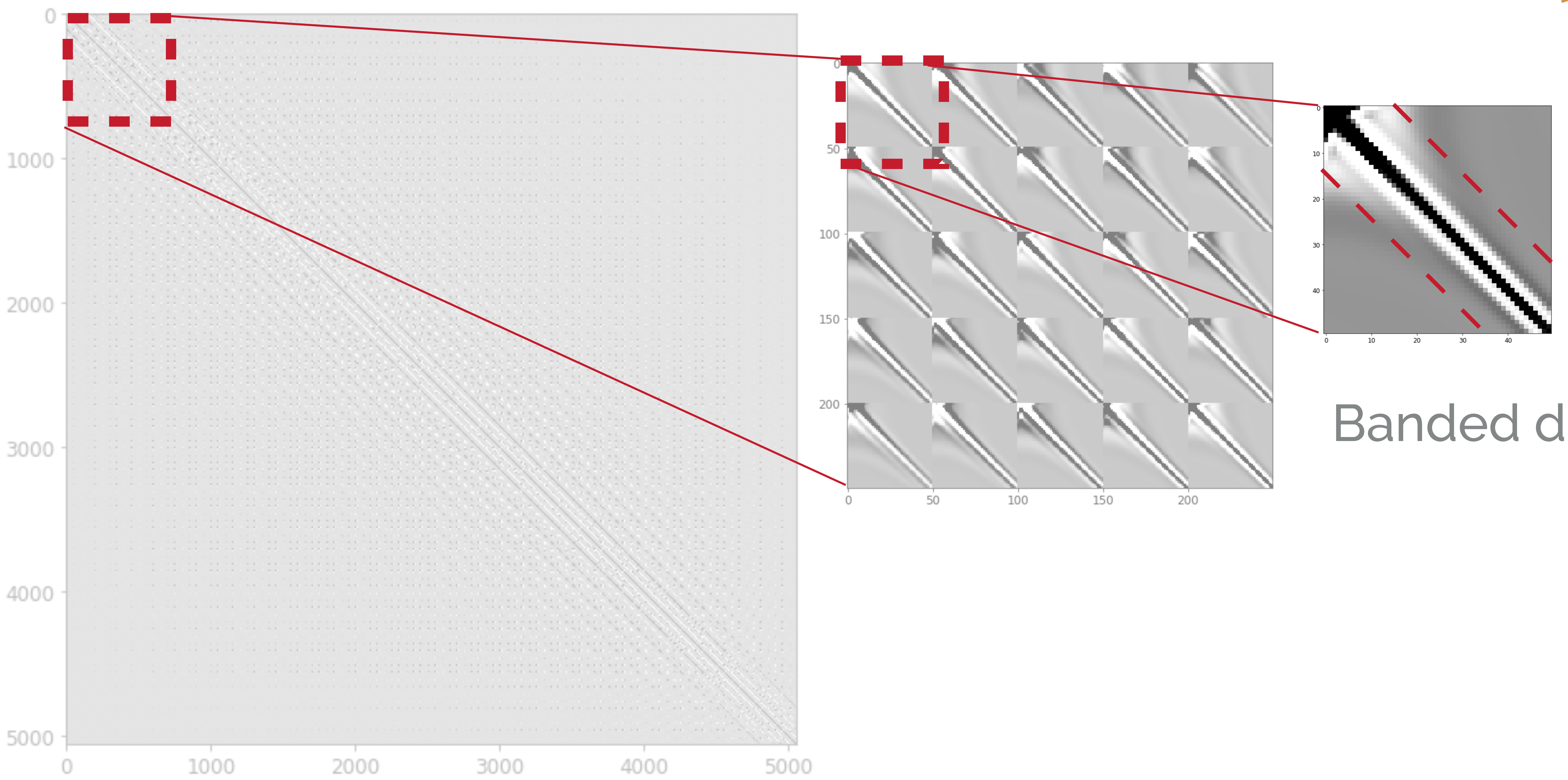



**Block dimensions:**

$n_z \times n_z$

**$n_x$ total blocks**

# Structure of the Hessian



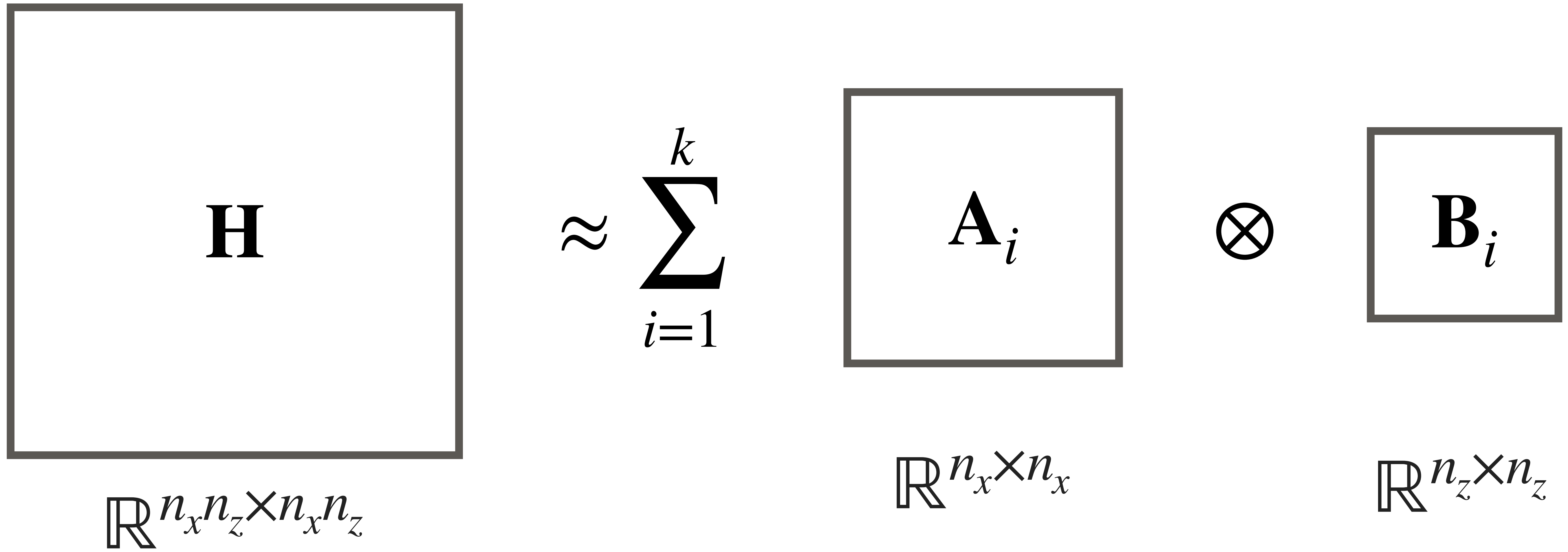

Banded diagonal

# Approximating the Hessian

Propose to approximate the Hessian using a superposition of Kronecker products

$$\mathbf{H} \approx \sum_{i=1}^{k} \mathbf{A}_i \otimes \mathbf{B}_i$$

Factors $\mathbf{A}_i$ and $\mathbf{B}_i$ are small matrices



# The Kronecker product

Suppose **A** and **B** are 2x2 matrices

$$\mathbf{A} = \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_1 & b_3 \\ b_2 & b_4 \end{bmatrix}$$

The Kronecker product is defined as

$$\mathbf{H} = \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix} \otimes \begin{bmatrix} b_1 & b_3 \\ b_2 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} & a_3 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} \\ a_2 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} & a_4 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}$$

# The Kronecker factors

Problem was addressed by Pitsianis (1997)

$$\mathbf{a}\,,\mathbf{b} = \arg\min_{\mathbf{a},\mathbf{b}} \|\tilde{\mathbf{H}} - \mathbf{b}\mathbf{a}^T\|_F^2$$

$$\tilde{\mathbf{H}} = \mathscr{R}(\mathbf{H}) \approx \sum_i^k \mathbf{b}_i \mathbf{a}_i^T$$ where $$\mathbf{a}_i = vec(\mathbf{A}_i) \qquad \mathbf{b}_i = vec(\mathbf{B}_i)$$

Factors are low-rank approximation of the *rearranged Hessian*

# The Kronecker factors

Problem was addressed by Pitsianis (1997)

$$\mathbf{a}\,,\mathbf{b} = \underset{\mathbf{a},\mathbf{b}}{\arg\,\min} \|\tilde{\mathbf{H}} - \mathbf{b}\mathbf{a}^T\|_F^2$$

$$\tilde{\mathbf{H}} = \mathscr{R}(\mathbf{H}) \approx \sum_i^k \mathbf{b}_i \mathbf{a}_i^T$$ where $$\mathbf{a}_i = vec(\mathbf{A}_i) \qquad \mathbf{b}_i = vec(\mathbf{B}_i)$$

Factors are low-rank approximation of the *rearranged Hessian*

↓

Estimate from SVD

# The rearranged Hessian

Rearrangement operator can be summarized as

$\mathbf{H}$ $\tilde{\mathbf{H}}$

$$\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix}$$

# The rearranged Hessian

Rearrangement operator can be summarized as

$$\mathbf{H} \qquad\qquad\qquad \tilde{\mathbf{H}}$$

$$\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \xrightarrow{\mathscr{R}} \begin{bmatrix} vec(\mathbf{H}_{11}) & vec(\mathbf{H}_{21}) & vec(\mathbf{H}_{12}) & vec(\mathbf{H}_{22}) \end{bmatrix}$$

# The rearranged Hessian

Rearrangement operator can be summarized as

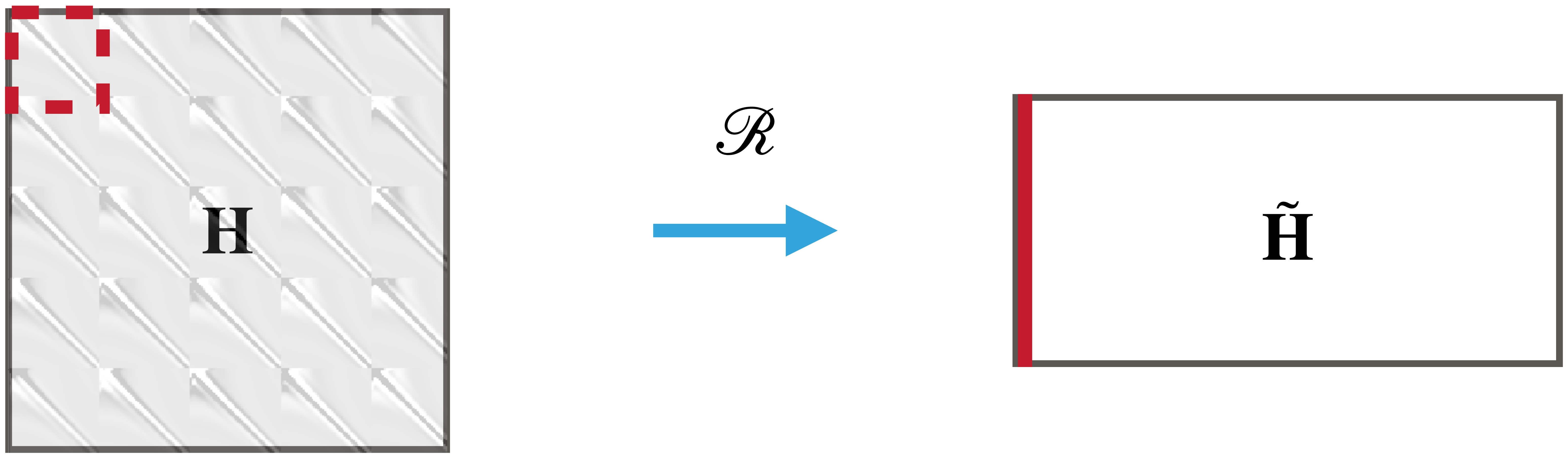
$$\mathbf{H} \qquad\qquad\qquad\qquad \tilde{\mathbf{H}}$$

$$\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \xrightarrow{\mathscr{R}} \begin{bmatrix} vec(\mathbf{H}_{11}) & vec(\mathbf{H}_{21}) & vec(\mathbf{H}_{12}) & vec(\mathbf{H}_{22}) \end{bmatrix}$$

$$\begin{bmatrix} a_1 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} & a_3 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} \\ a_2 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} & a_4 \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} \end{bmatrix} \qquad \begin{bmatrix} a_1b_1 & a_2b_1 & a_3b_1 & a_4b_1 \\ a_1b_2 & a_2b_2 & a_3b_2 & a_4b_2 \\ a_1b_3 & a_2b_3 & a_3b_3 & a_4b_3 \\ a_1b_4 & a_2b_4 & a_3b_4 & a_4b_4 \end{bmatrix}.$$

# The rearranged Hessian

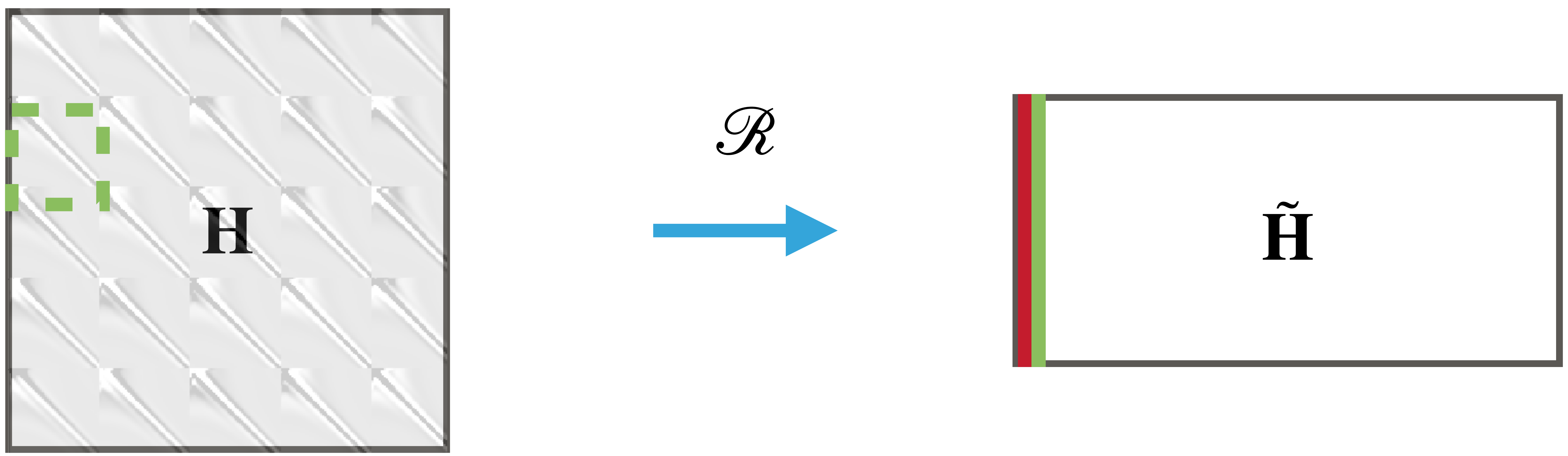
Rearrangement operator can be summarized as




$\mathbb{R}^{n_x n_z \times n_x n_z}$ $\qquad$ $\mathbb{R}^{n_z^2 \times n_x^2}$

# The rearranged Hessian

Rearrangement operator can be summarized as

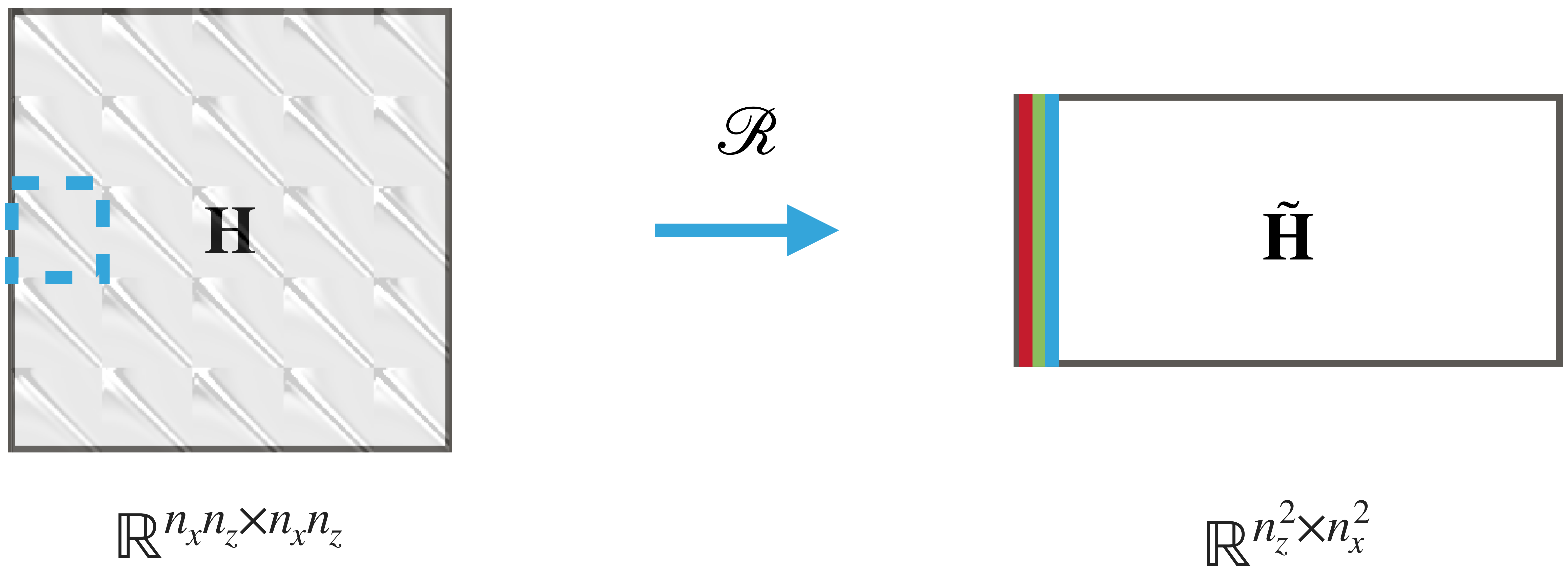



$\mathbb{R}^{n_x n_z \times n_x n_z}$

$\mathbb{R}^{n_z^2 \times n_x^2}$

# The rearranged Hessian

Rearrangement operator can be summarized as



# Hessian-vector products

The Kronecker product possesses a useful property

$$(\mathbf{A} \otimes \mathbf{B})\mathbf{m} = vec(\mathbf{B}\mathbf{M}\mathbf{A}^T), \quad \mathbf{m} = vec(\mathbf{M})$$

# Hessian-vector products

The Kronecker product possesses a useful property

$$\overset{\mathbb{R}^{n_x n_z \times n_x n_z}}{\boxed{\mathbf{A}\otimes\mathbf{B}}}\ \overset{\mathbb{R}^{n_x n_z \times 1}}{\boxed{\mathbf{m}}} = vec(\overset{\mathbb{R}^{n_z \times n_z}}{\boxed{\mathbf{B}}}\ \overset{\mathbb{R}^{n_z \times n_x}}{\boxed{\mathbf{M}}}\ \overset{\mathbb{R}^{n_x \times n_x}}{\boxed{\mathbf{A}^T}})$$

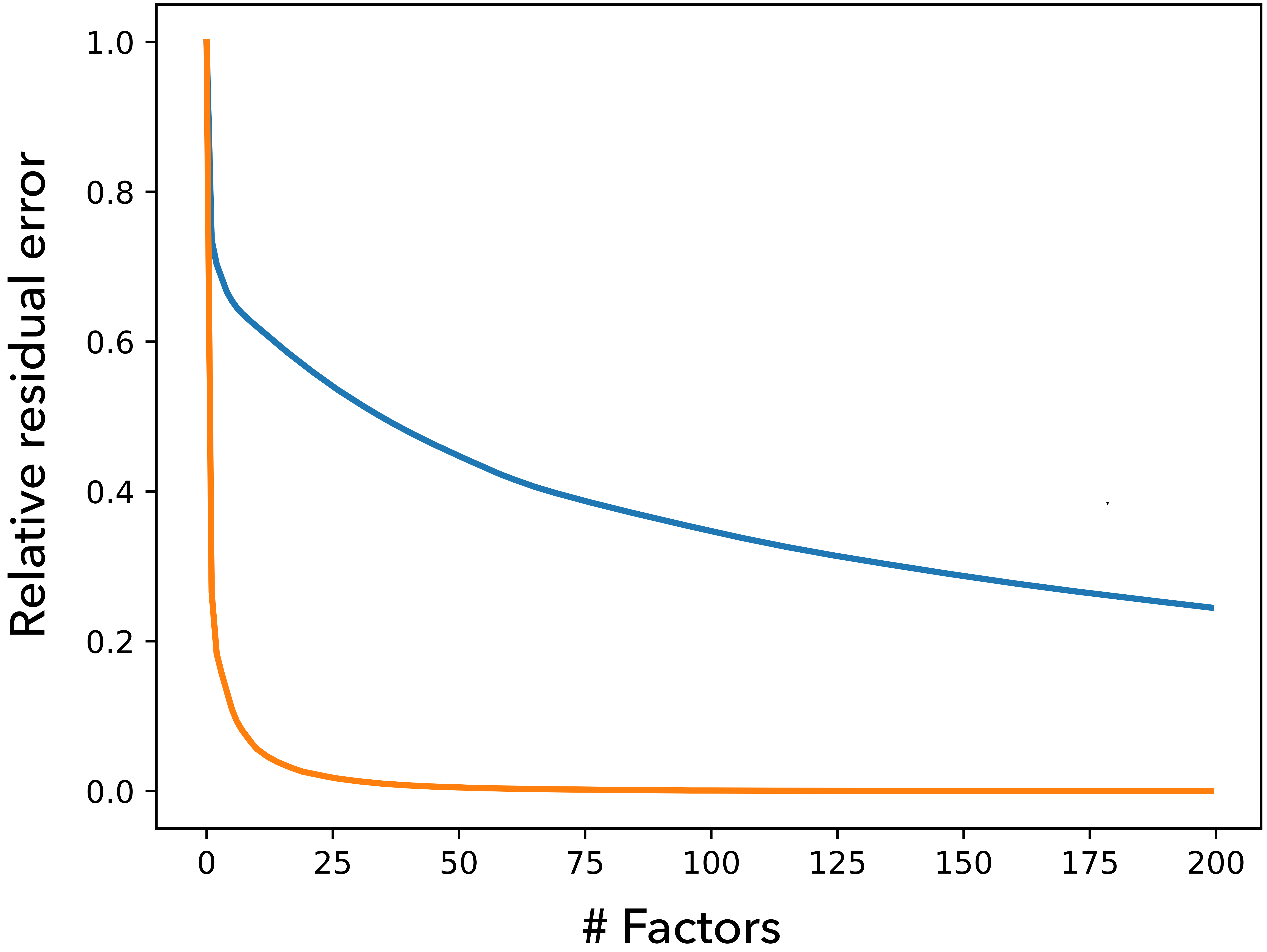
# Hessian-vector products

This allows a Hessian-vector product to be approximated as

$$\mathbf{Hm} \approx \sum_{i=1}^{k} vec(\mathbf{B}_i \mathbf{M} \mathbf{A}_i^T) \quad \longrightarrow \quad \sum_{i=1}^{k} vec\Big(\overset{\mathbb{R}^{n_z \times n_z}}{\boxed{\mathbf{B}_i}}\ \overset{\mathbb{R}^{n_z \times n_x}}{\boxed{\quad \mathbf{M} \quad}}\ \overset{\mathbb{R}^{n_x \times n_x}}{\boxed{\quad \mathbf{A}_i^T \quad}}\Big)$$

**Can approximate Hessian-vector products using operations with small matrices**

# Approximation error



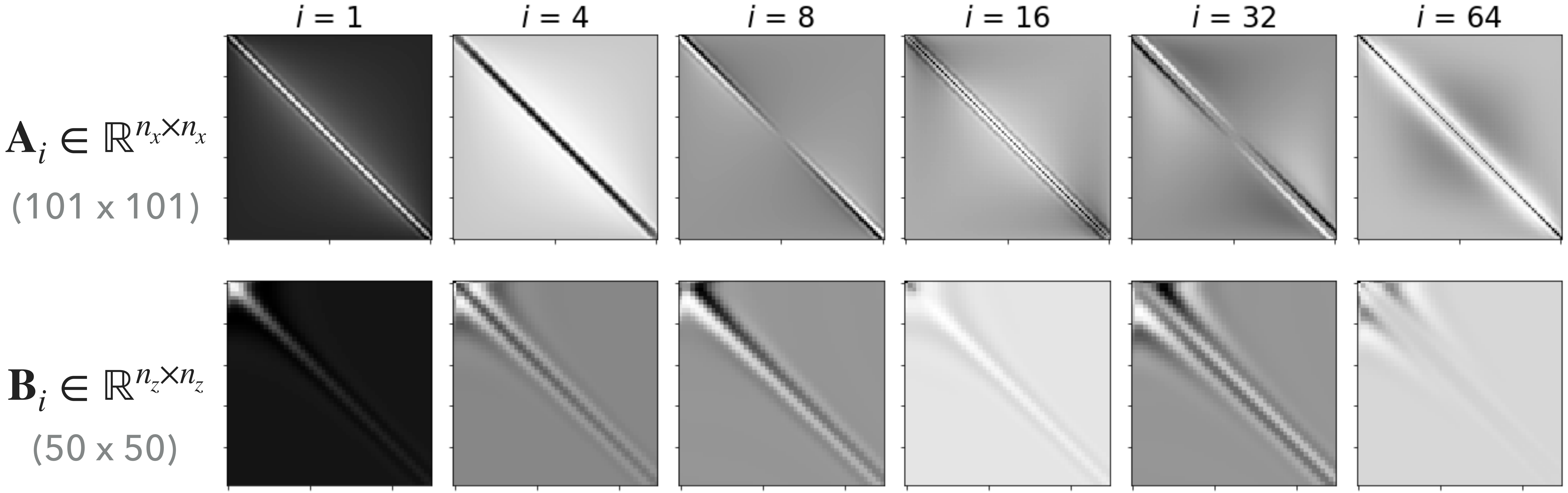

Error in low-rank approximation

$$\frac{\|\mathbf{H} - \sum_i^k \lambda_i \mathbf{u}_i \mathbf{v}_i^T\|_F}{\|\mathbf{H}\|_F}$$

Error in Kronecker approximation

$$\frac{\|\mathbf{H} - \sum_i^k \mathbf{A}_i \otimes \mathbf{B}_i\|_F}{\|\mathbf{H}\|_F}$$

# Kronecker factors



# Estimating the factors

**Small problems:** Direct SVD of rearranged Hessian

**Medium/Large problems:** Estimate factors using low-rank matrix completion of rearranged Hessian

$$\tilde{\mathbf{A}}, \tilde{\mathbf{B}} = \underset{\tilde{\mathbf{A}} \in \mathbb{R}^{n_z{}^2 \times k},\ \tilde{\mathbf{B}} \in \mathbb{R}^{n_x{}^2 \times k}}{\arg\min} \; ||P_\Omega(\tilde{\mathbf{H}} - \tilde{\mathbf{B}}\tilde{\mathbf{A}}^T)||_F^2, \qquad \tilde{\mathbf{A}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$$

Compute samples preferentially according to the structure

# Estimating the factors

**Small problems:** Direct SVD of rearranged Hessian

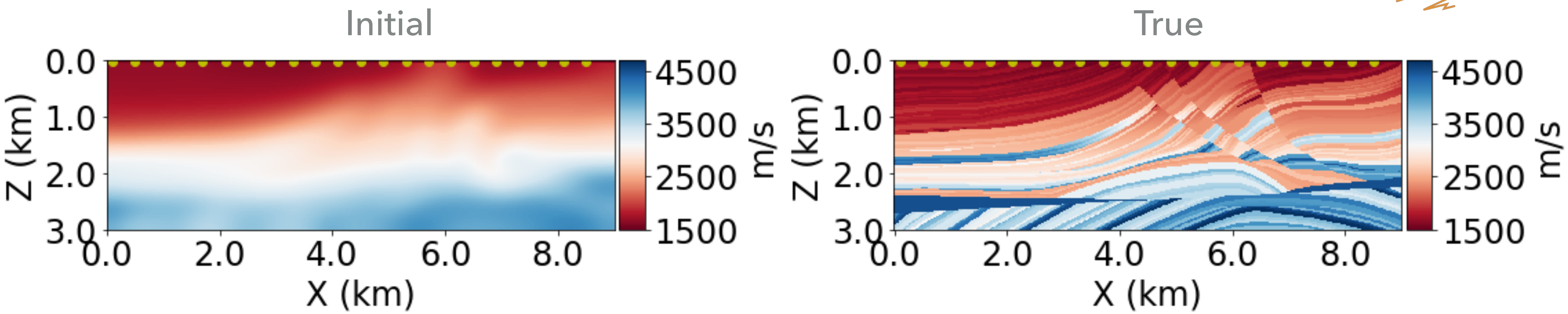
**Medium/Large problems:** Estimate factors using low-rank matrix completion of rearranged Hessian

$$\tilde{\mathbf{A}}, \tilde{\mathbf{B}} = \underset{\tilde{\mathbf{A}} \in \mathbb{R}^{n_z{}^2 \times k},\ \tilde{\mathbf{B}} \in \mathbb{R}^{n_x{}^2 \times k}}{\arg\min} ||P_\Omega(\tilde{\mathbf{H}} - \tilde{\mathbf{B}}\tilde{\mathbf{A}}^T)||_F^2, \qquad \tilde{\mathbf{A}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$$

Requires PDE solves - Limit cost by using receiver Green's functions ($N_s$ + $N_r$)

# Numerical Experiments

# Marmousi model

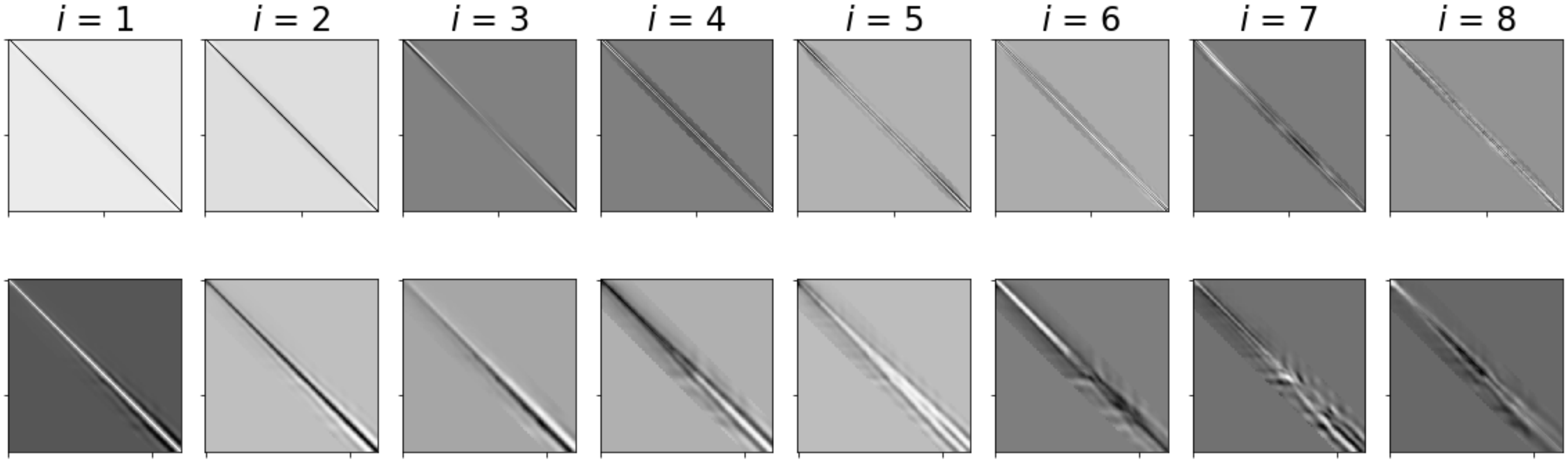



- Time-domain acoustic modelling (900 x 300)
- 22 sources, 225 receivers distributed at surface
- Multi-scale inversion (3-5Hz, -7Hz, -9Hz)
- 30 preconditioned NLCG iterations per frequency band

# Kronecker factors

$\mathbf{A}_i \in \mathbb{R}^{n_x \times n_x}$
(900 x 900)

$\mathbf{B}_i \in \mathbb{R}^{n_z \times n_z}$
(300 x 300)

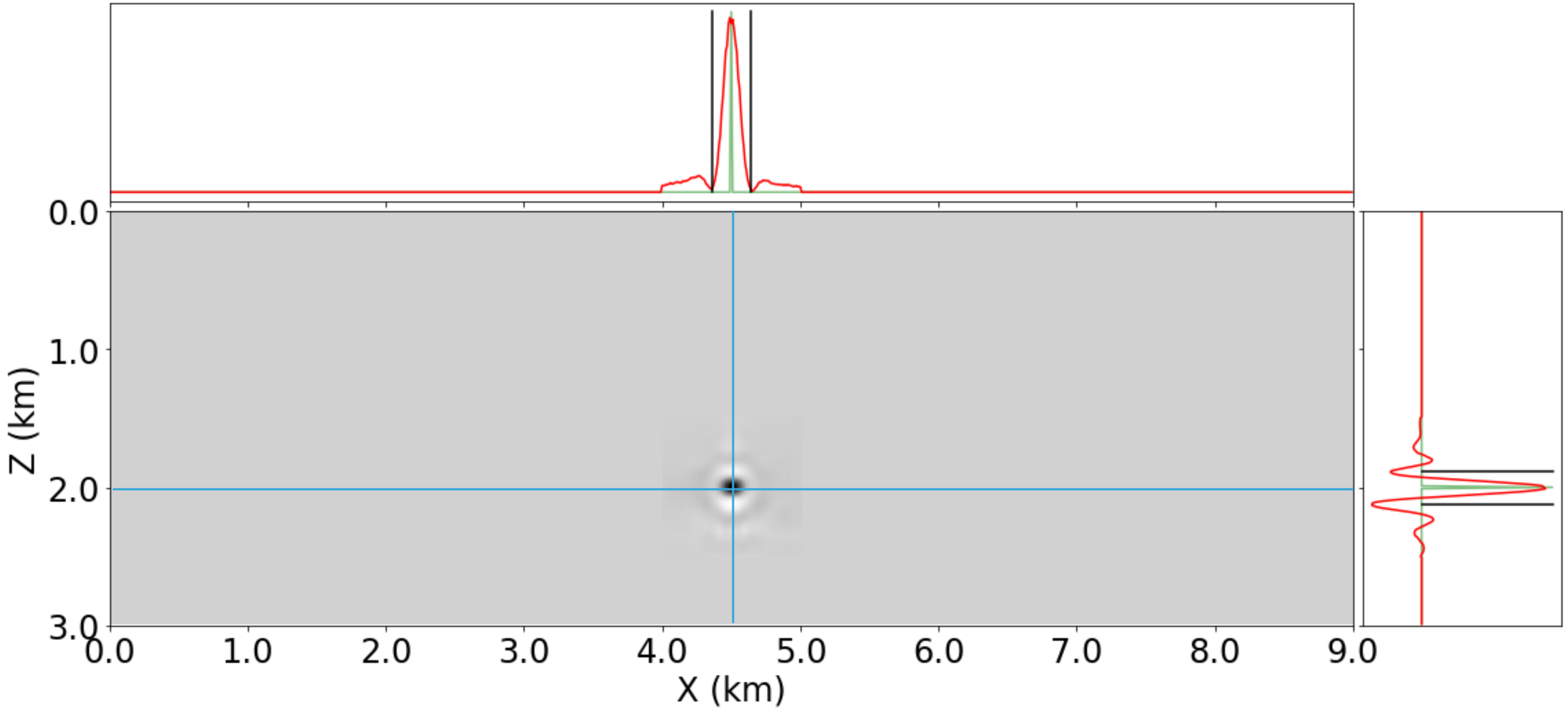



Estimated from 0.2% of the total number of elements in the Hessian ($7.29 \times 10^{10}$)

# Local Resolution Analysis

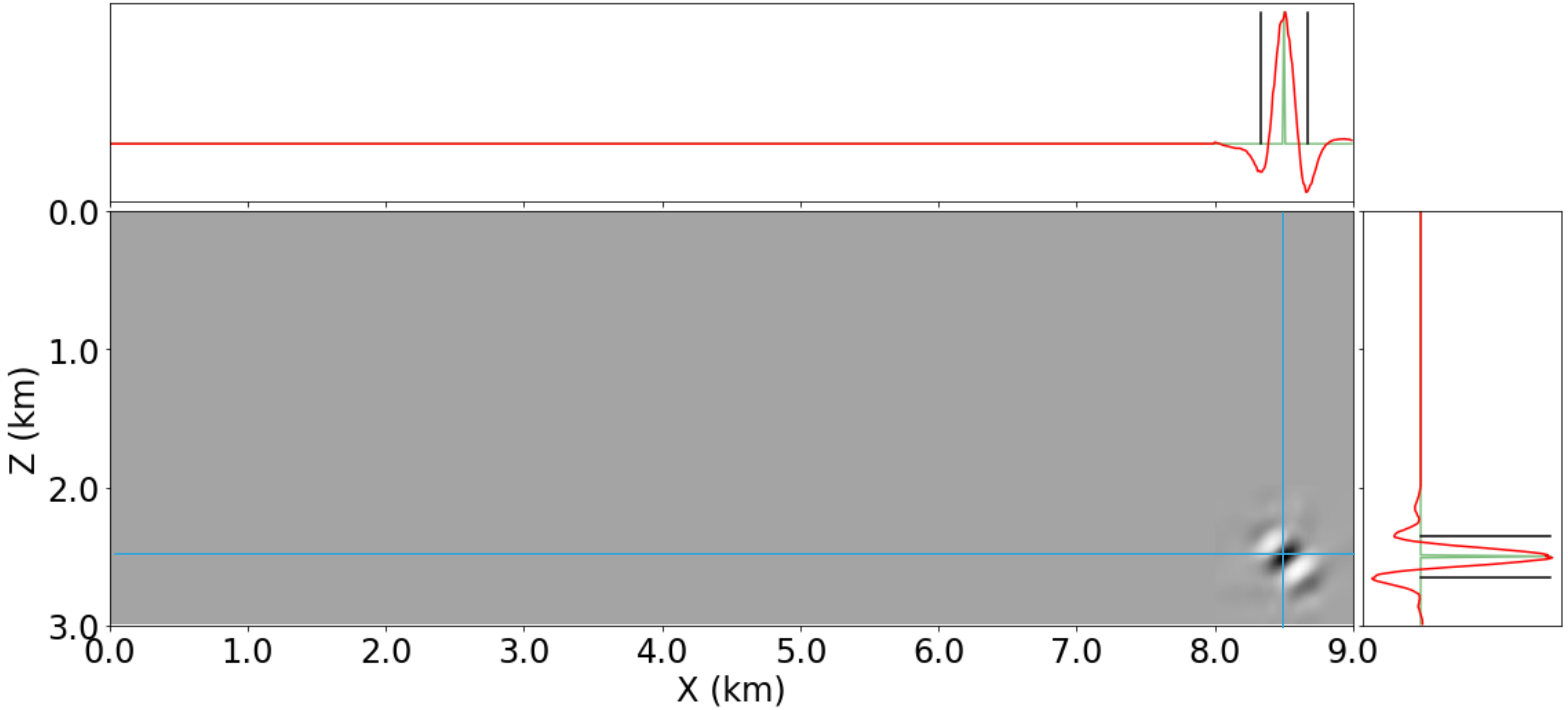
# Probing the subsurface

Estimate local resolution lengths by applying the Hessian to point scatterers at various locations



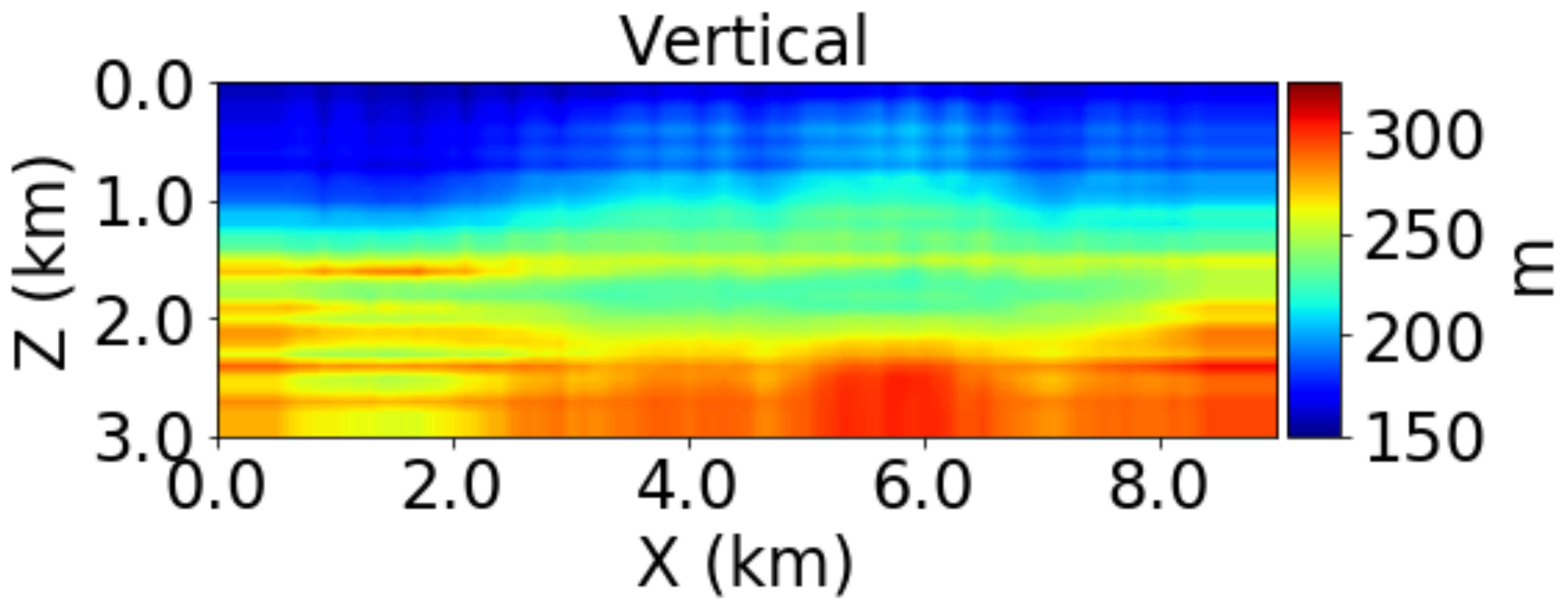
# Probing the subsurface

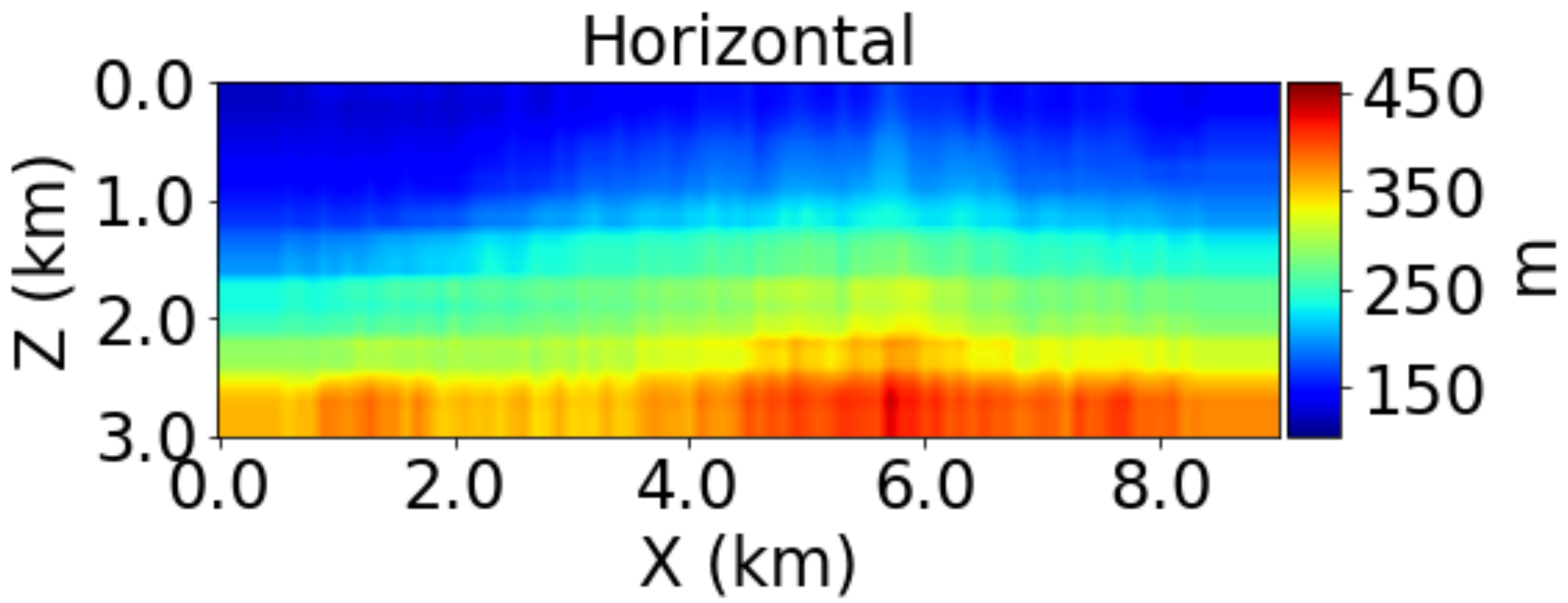
Vertical and horizontal resolution lengths become less meaningful when scatterers are not isotropically smeared



# Resolution lengths

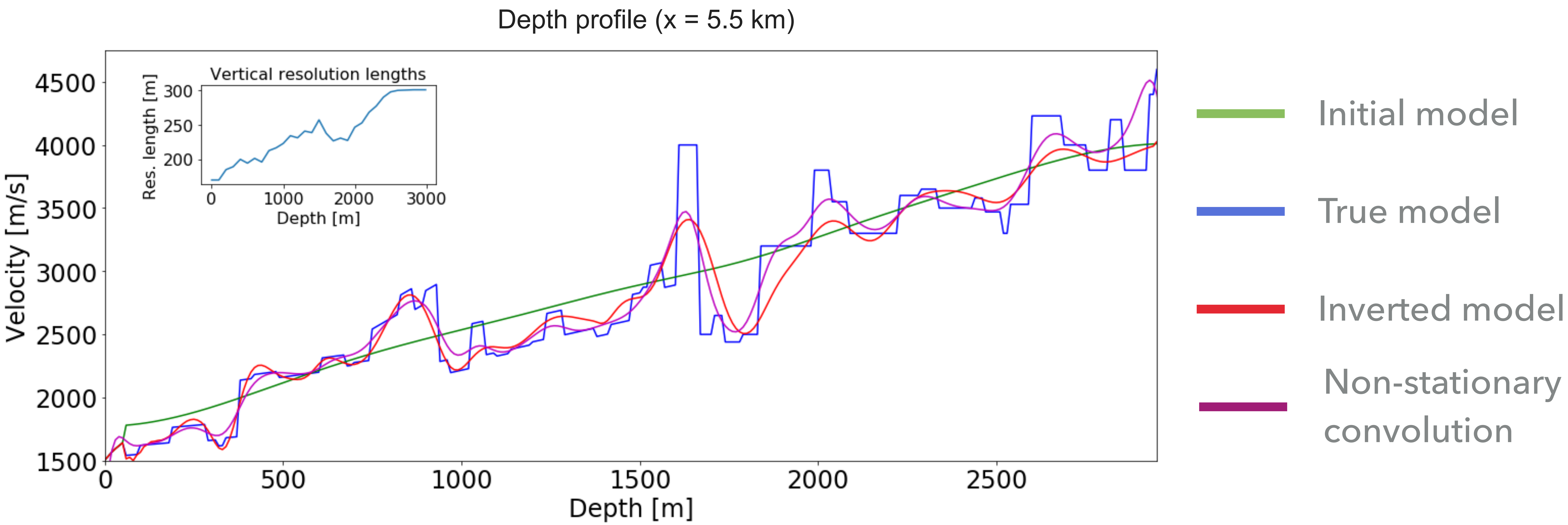
Probe the subsurface to generate maps of horizontal and vertical resolution lengths (interpolated, queried every 10 m).



# Non-stationary convolution

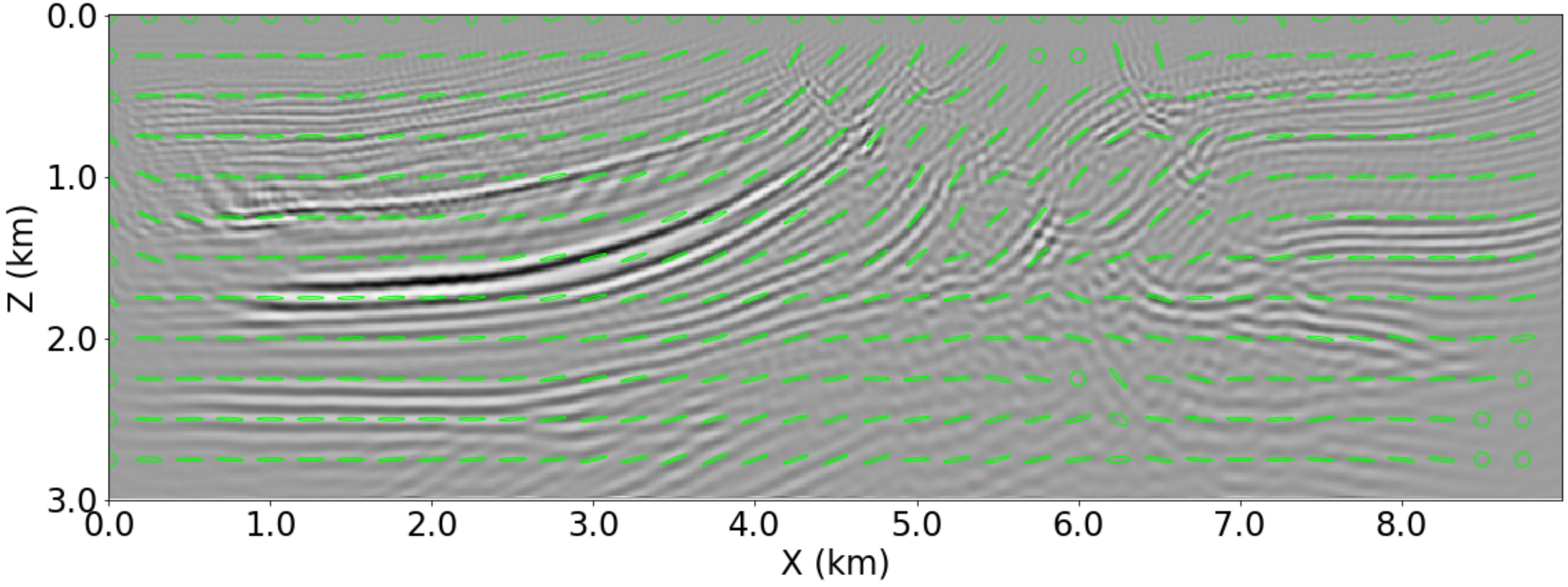
Perform 1D non-stationary convolution on the true model with a series of Gaussian filters parametrized by the variable resolution lengths



# Linearized Bayesian Inversion

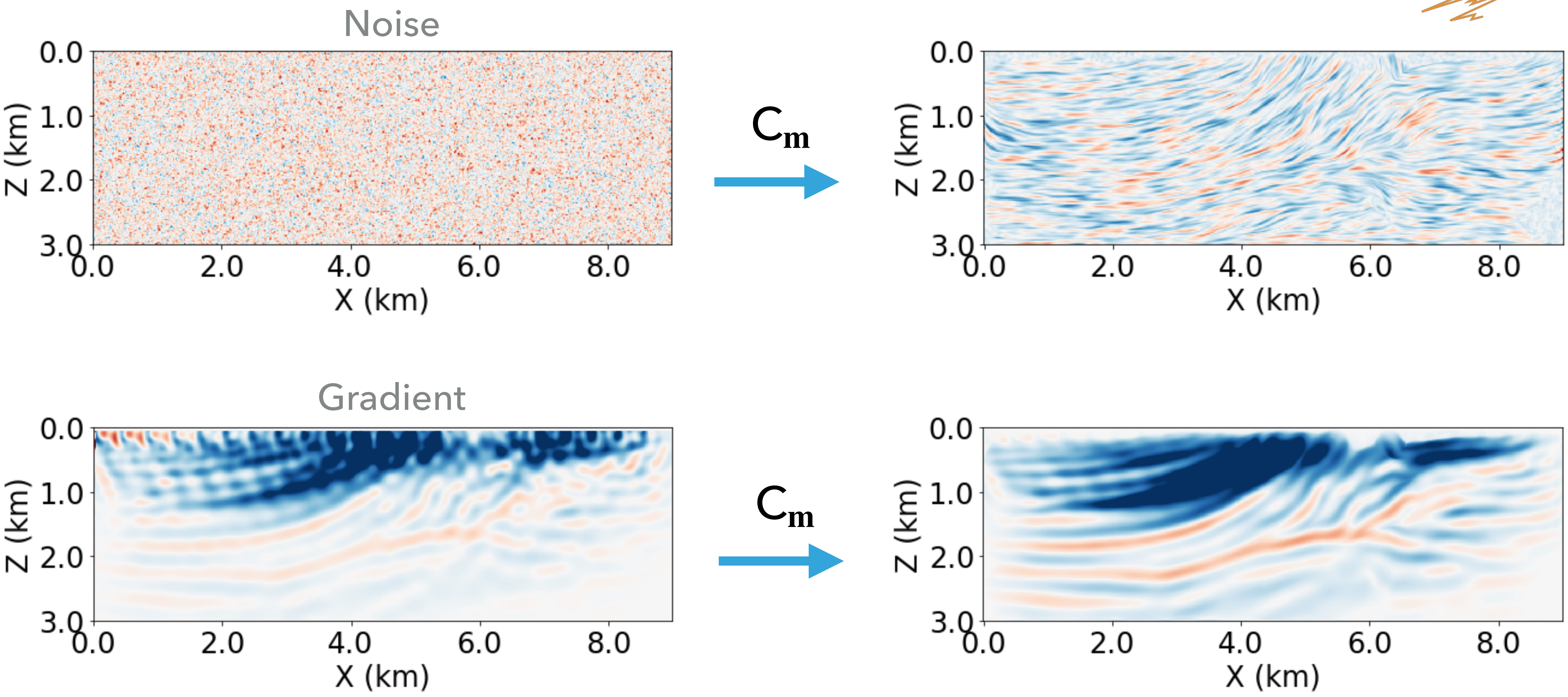
# Prior covariance - $C_m$

Use an anisotropic Matern covariance (Hale, 2014)



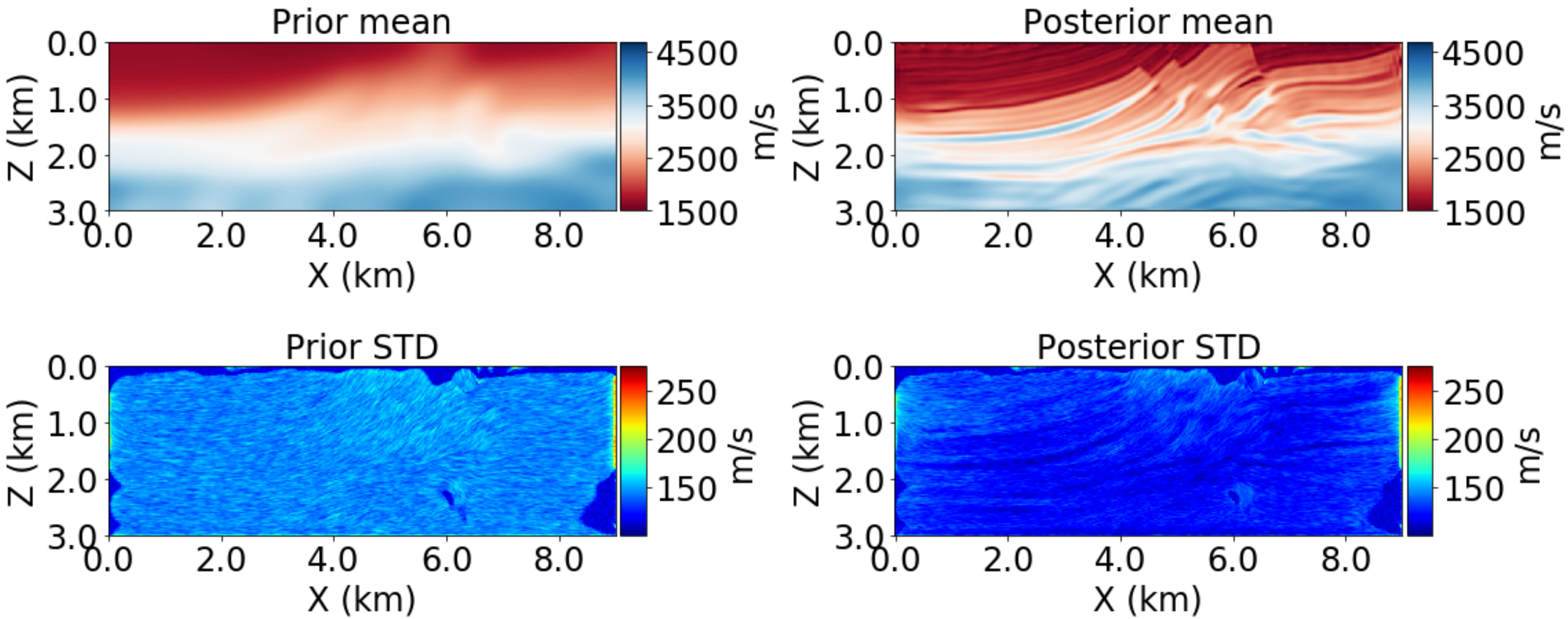

Allows for structure-oriented smoothing informed by a seismic image

# Shaping covariance (Hale, 2014)

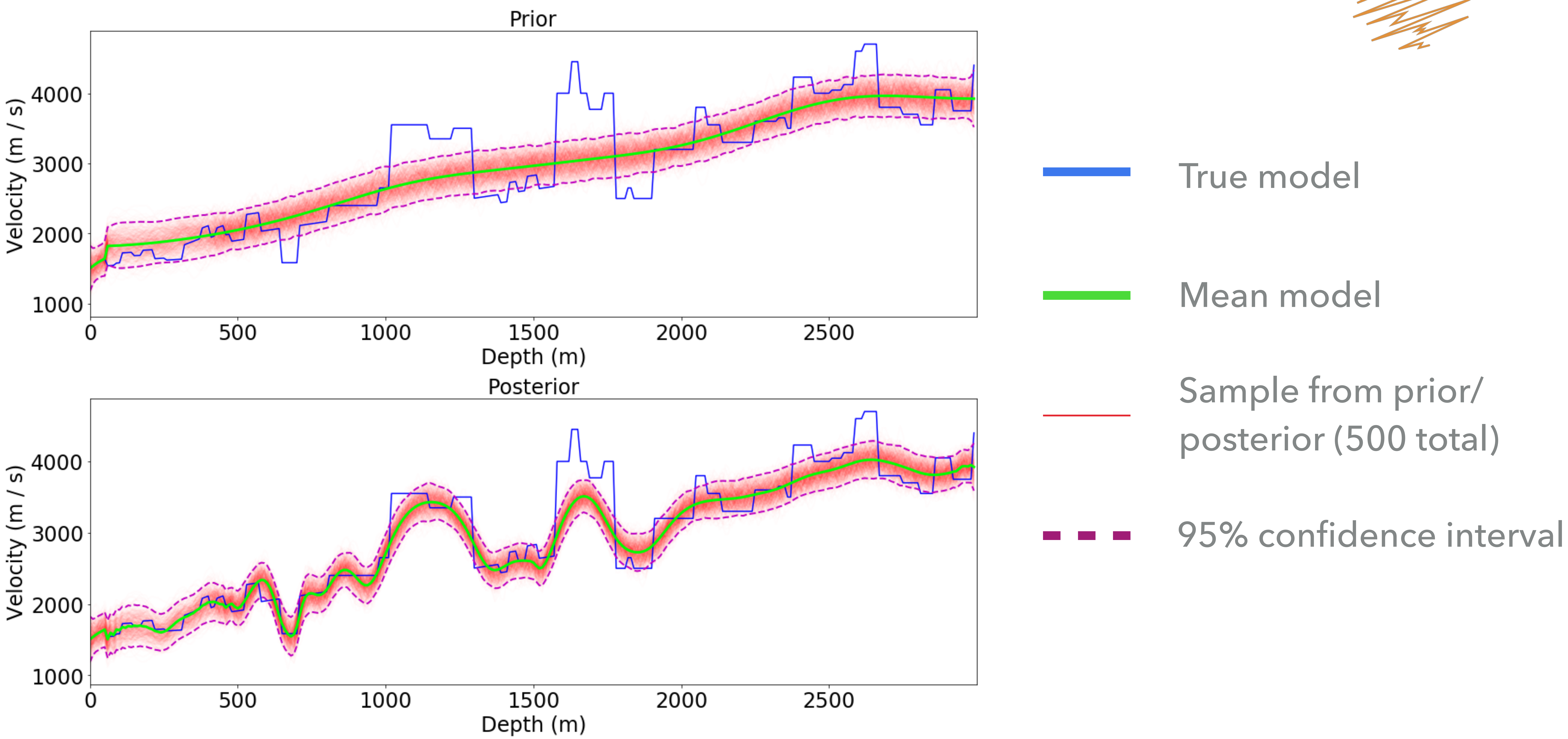


# Comparing distributions

Sample means and standard deviations computed over 500 random samples



# Depth profile comparison (x = 6.4 km)



# Conclusions

- Propose a novel **factorization of the Hessian** as a **superposition of Kronecker products**
- Use Kronekcer factorization for **efficient Hessian-vector products**
- Presented two forms of resolution/uncertainty analysis using fast Hessian-vector products
- Extensions to 3D are challenging as they involve Tensor-completion (via Canonical/Parafac decomposition)

# Acknowledgements

We would like to thank the sponsors of the Signal analysis and Imaging Group at the University of Alberta for their continued support.

This research has been funded by the NSERC Alexander Graham Bell Canada Graduate scholarship and Alberta Innovates Technology Futures scholarship.

# References


- N.P. Pitsianis, The Kronecker product in approximation, fast transform generation, Ph.D. Thesis, Department of Computer Science, Cornell University, 1997.
- Tarantola, A., 2005, Inverse problem theory and method for model parameter estimation: SIAM.
- Plessix, R., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, 167, 495–503.
- Bui-Thanh, T., O. Ghattas, J. Martin, and G. Stadler, 2013, A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion: SIAM Journal on Scientific Computing, 35, A2494–A2523.
- Hale, D., 2014, Implementing an anisotropic and spatially varying Matérn model covariance with smoothing filters: CWP Report, 815, 281–290.
- Fichtner, A., and T. van Leeuwen, 2015, Resolution analysis by random probing: Journal of Geophysical Research, 120, 5549–5573.
- Zhu H, Li S, Fomel S, Stadler G, Ghattas O (2016) A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration. Geophysics 81(5):R307–R323.