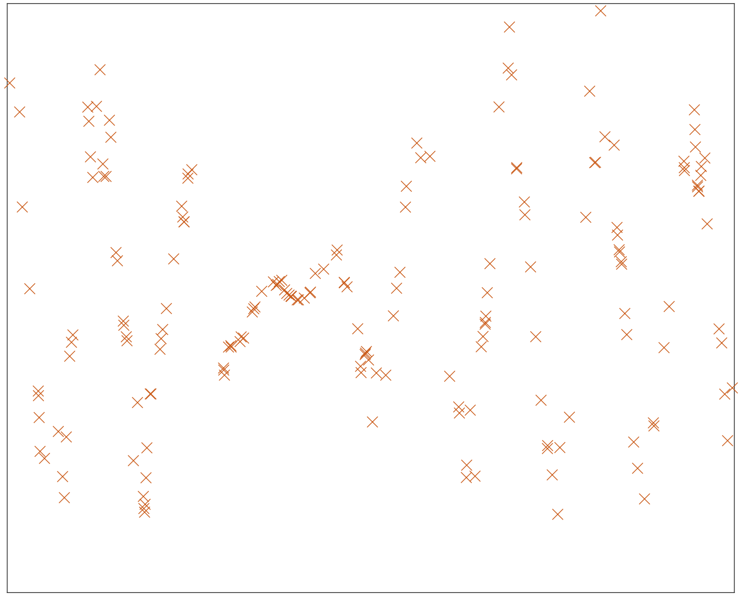
Iteratively Reweighted Least Squares Recovery on Tensor Tree Networks

Sebastian Krämer

Teaser — Exemplary Application

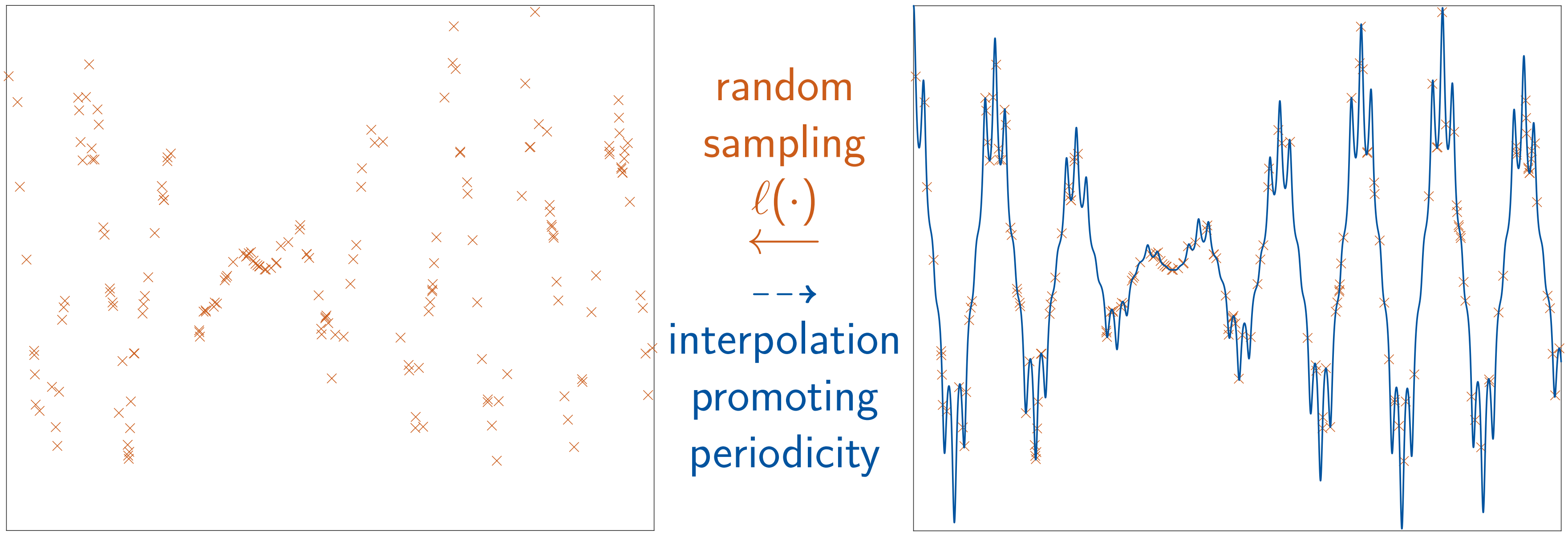


random
sampling
$\ell(\cdot)$
$\longleftarrow$

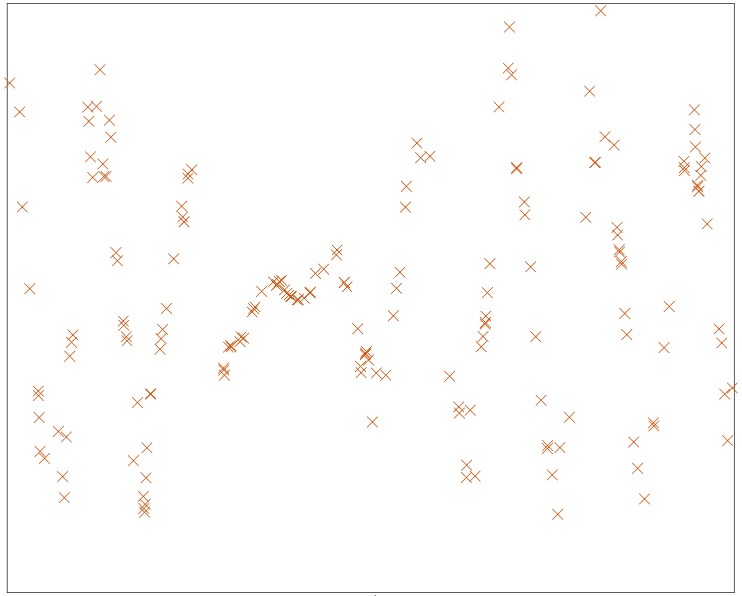
$\dashrightarrow$
interpolation
promoting
periodicity

?

Teaser — Exemplary Application

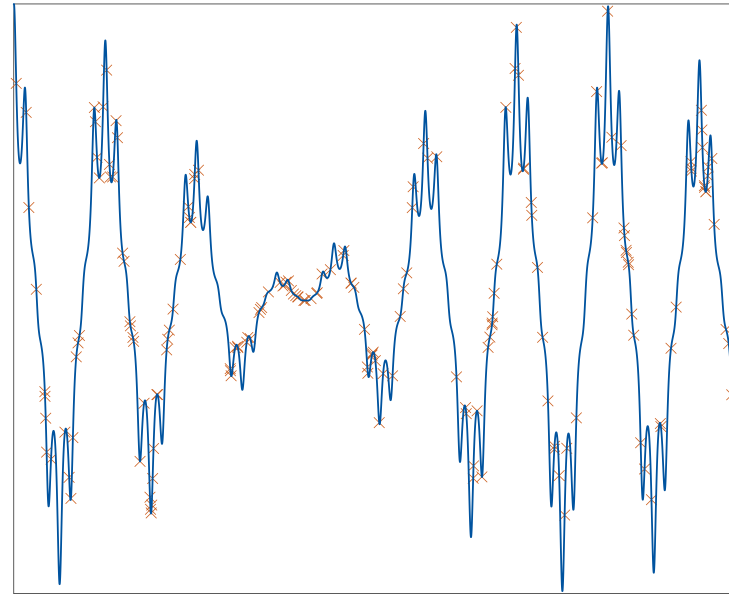


Teaser — Exemplary Application



random sampling $\ell(\cdot)$ $\longleftarrow$

$\dashrightarrow$ interpolation promoting periodicity



affine set $\downarrow$

$\{g \in \mathbb{R}^{2^d} \mid \ell(g) = y\}$

quantization $\downarrow$

$\{X \in \mathbb{R}^{2\times\dots\times 2} \mid \mathcal{L}(X) = y\}$

$\xrightarrow{\text{lowest TT ranks}}$

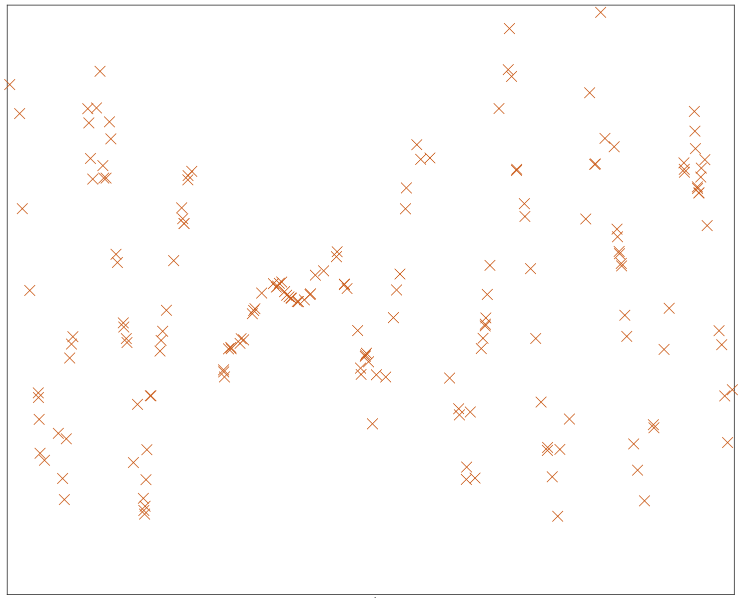
X^*

$\uparrow$ vectorization

g^*

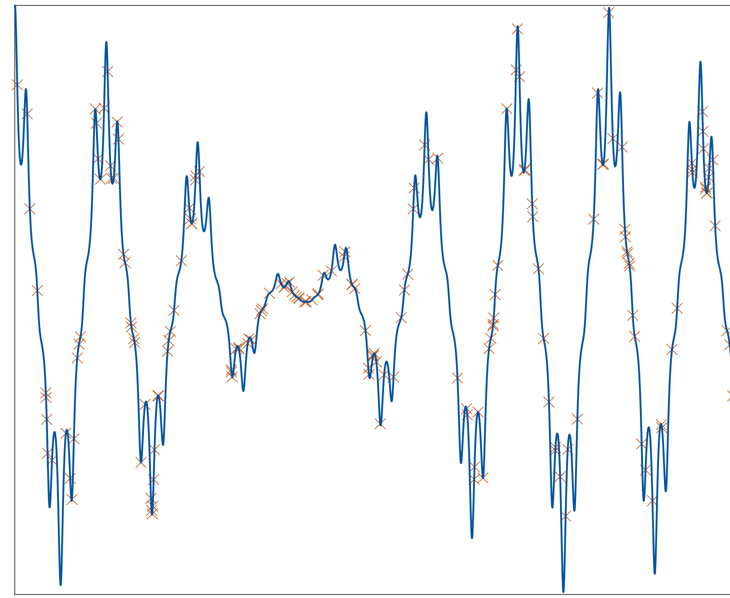
$\uparrow$ approximation

Teaser — Exemplary Application



random sampling $\ell(\cdot)$ $\longleftarrow$

$\dashrightarrow$ interpolation promoting periodicity



affine set $\downarrow$

$\{g \in \mathbb{R}^{2^d} \mid \ell(g) = y\}$

quantization $\downarrow$

$\{X \in \mathbb{R}^{2\times\dots\times 2} \mid \mathcal{L}(X) = y\}$ $\xrightarrow{\text{lowest TT ranks}}$ X^*

$\uparrow$ vectorization

g^*

$\uparrow$ approximation

Require:

- rank *minimizing* tensor completion applicable in high dimensions

SIMPLE FIRST: MATRICES

Given:

- (surjective) linear operator $\mathcal{L} : \mathbb{R}^{n \times m} \to \mathbb{R}^{\ell}$
- measurements $y \in \mathbb{R}^{\ell}$

Definition (affine rank minimization problem)

$$\text{find} \quad X^* \in \operatorname*{argmin}_{X \in \mathbb{R}^{n \times m}} \operatorname{rank}(X) \quad \text{subject to} \quad \mathcal{L}(X) = y$$

Among all matrices satisfying given linear equations, find the one with lowest rank.

Affine because the preset $\mathcal{L}^{-1}(y) = \{X \in \mathbb{R}^{n \times m} \mid \mathcal{L}(X) = y\}$ is affine.

Affine rank minimization — Asymptotic log-det minimization

Approach via minimization of:

Definition (log-det family[e.g. Fazel et al. '03])

$$f_\gamma(X) := \log\det(XX^T + \gamma I_n), \quad \gamma > 0$$

Affine rank minimization — Asymptotic log-det minimization

Approach via minimization of:

Definition (log-det family[e.g. Fazel et al. '03])

$$f_\gamma(X) := \log\det(XX^T + \gamma I_n), \quad \gamma > 0$$

Smoothed Schatten-p-functions:

$$S_{\gamma,p}(X) := \sum_{j=1}^{n} (\sigma_j^2(X) + \gamma)^{\frac{p}{2}}, \quad 0 < p \leq 1$$

- $p = 1$: (smoothed) convex nuclear norm:
$$S_{0,1}(X) = \|X\|_*$$
- "$p = 0$": log-det functions:
$$\lim_{p \searrow 0} \frac{S_{\gamma,p}(X) - n}{p} = \frac{1}{2} f_\gamma(X)$$

Iteratively Reweighted Least Squares Recovery on Tensor Networks — Sebastian Krämer, kraemer@igpm.rwth-aachen.de — RWTH Aachen University — 17.05.24

Approach via minimization of:

Definition (log-det family[e.g. Fazel et al. '03])

$$f_\gamma(X) := \log\det(XX^T + \gamma I_n), \quad \gamma > 0$$

Smoothed Schatten-p-functions:

$$S_{\gamma,p}(X) := \sum_{j=1}^{n} (\sigma_j^2(X) + \gamma)^{\frac{p}{2}}, \quad 0 < p \leq 1$$

- $p = 1$: (smoothed) convex nuclear norm:

$$S_{0,1}(X) = \|X\|_*$$

- "$p = 0$": log-det functions:

$$\lim_{p \searrow 0} \frac{S_{\gamma,p}(X) - n}{p} = \frac{1}{2} f_\gamma(X) \quad \text{but} \quad f_0(X) = \min \Leftrightarrow X \text{ singular}$$

Example (matrix rank minimization)

$$X_\delta(a) := \begin{pmatrix} \delta & \sqrt{\delta} \\ \sqrt{\delta} & a \end{pmatrix}, \quad \text{seek } a^* = \begin{cases} 1, & \delta > 0 \quad \text{(rank 1)} \\ 0, & \delta = 0 \quad \text{(rank 0)} \end{cases}$$

Example (matrix rank minimization)

$$X_\delta(a) := \begin{pmatrix} \delta & \sqrt{\delta} \\ \sqrt{\delta} & a \end{pmatrix}, \quad \text{seek } a^* = \begin{cases} 1, & \delta > 0 \quad (\text{rank } 1) \\ 0, & \delta = 0 \quad (\text{rank } 0) \end{cases}$$

Minimize $S_{0,p}$ (Schatten-p, $p > 0$) or $f_\gamma(X) = \log\det(XX^T + \gamma I_n)$:

Affine rank minimization — Asymptotic log-det minimization

Example (matrix rank minimization)

$$X_\delta(a) := \begin{pmatrix} \delta & \sqrt{\delta} \\ \sqrt{\delta} & a \end{pmatrix}, \quad \text{seek } a^* = \begin{cases} 1, & \delta > 0 \quad \text{(rank 1)} \\ 0, & \delta = 0 \quad \text{(rank 0)} \end{cases}$$

Minimize $S_{0,p}$ (Schatten-p, $p > 0$) or $f_\gamma(X) = \log\det(XX^T + \gamma I_n)$:

- $p > 0$, $\delta > 0$:

$$\delta < \frac{1}{\sqrt[p]{4} - 1} \quad \Rightarrow \quad S_{0,p}(X_\delta(0)) < S_{0,p}(X_\delta(1)) \qquad ✗$$

Example (matrix rank minimization)

$$X_\delta(a) := \begin{pmatrix} \delta & \sqrt{\delta} \\ \sqrt{\delta} & a \end{pmatrix}, \quad \text{seek } a^* = \begin{cases} 1, & \delta > 0 \quad (\text{rank } 1) \\ 0, & \delta = 0 \quad (\text{rank } 0) \end{cases}$$

Minimize $S_{0,p}$ (Schatten-p, $p > 0$) or $f_\gamma(X) = \log\det(XX^T + \gamma I_n)$:

- $p > 0$, $\delta > 0$:

$$\delta < \frac{1}{\sqrt[p]{4} - 1} \quad \Rightarrow \quad S_{0,p}(X_\delta(0)) < S_{0,p}(X_\delta(1)) \qquad ✗$$

- $p = 0$, $\delta = 0$, $\gamma = 0$:

$$X_0(a) = \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix}, \quad f_0(X_0(a)) \equiv \text{const} \qquad ✗$$

Example (matrix rank minimization)

$$X_\delta(a) := \begin{pmatrix} \delta & \sqrt{\delta} \\ \sqrt{\delta} & a \end{pmatrix}, \quad \text{seek } a^* = \begin{cases} 1, & \delta > 0 \quad \text{(rank 1)} \\ 0, & \delta = 0 \quad \text{(rank 0)} \end{cases}$$

Minimize $S_{0,p}$ (Schatten-p, $p > 0$) or $f_\gamma(X) = \log\det(XX^T + \gamma I_n)$:

- $p > 0$, $\delta > 0$:

$$\delta < \frac{1}{\sqrt[p]{4} - 1} \quad \Rightarrow \quad S_{0,p}(X_\delta(0)) < S_{0,p}(X_\delta(1)) \qquad ✗$$

- $p = 0$, $\delta = 0$, $\gamma = 0$:

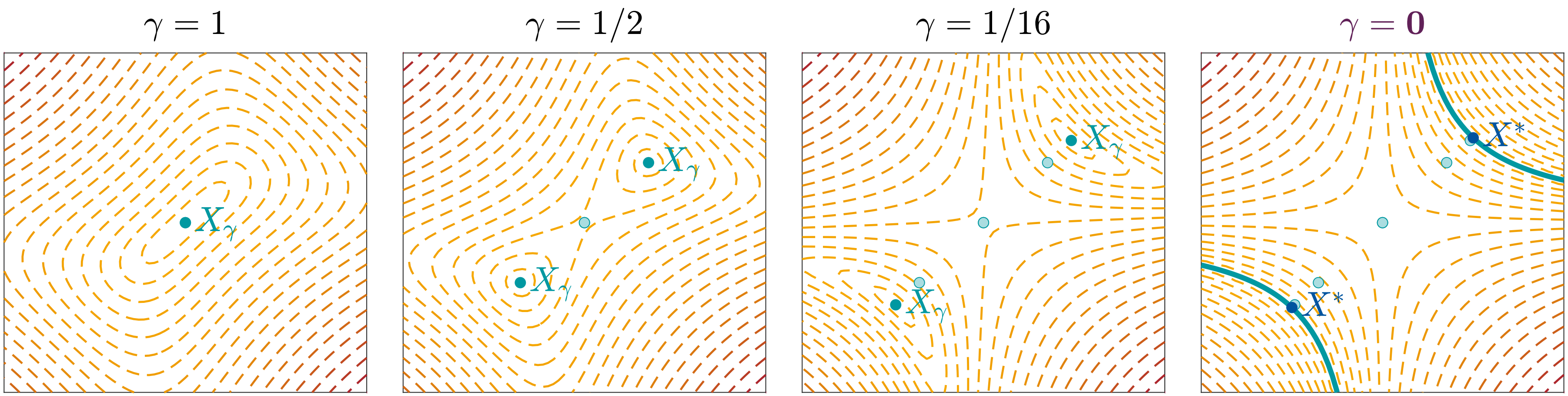
$$X_0(a) = \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix}, \quad f_0(X_0(a)) \equiv \mathrm{const} \qquad ✗$$

- *asymptotic minimization*:

$$\lim_{\gamma \searrow 0} \operatorname{argmin}_{a \in \mathbb{R}} f_\gamma(X_\delta(a)) = \lim_{\gamma \searrow 0} \frac{\delta^2}{\delta^2 + \gamma} = \begin{cases} 1, & \delta > 0 \qquad ✓ \\ 0, & \delta = 0 \qquad ✓ \end{cases}$$

Search limits X^* of (global) minimizers X_γ for $\gamma \searrow 0$:

$$\mathcal{X}^* := \left\{ X^* \mid \exists (X_\gamma)_{\gamma>0}, \quad X^* = \lim_{\gamma \searrow 0} X_\gamma, \; X_\gamma \in \operatorname*{argmin}_{X \in \mathcal{L}^{-1}(y)} f_\gamma(X) \right\}$$



Affine rank minimization — Asymptotic log-det minimization

Search limits X^* of (global) minimizers X_γ for $\gamma \searrow 0$:

$$\mathcal{X}^* := \left\{ X^* \mid \exists (X_\gamma)_{\gamma>0}, \quad X^* = \lim_{\gamma \searrow 0} X_\gamma, \ X_\gamma \in \operatorname*{argmin}_{X \in \mathcal{L}^{-1}(y)} f_\gamma(X) \right\}$$

Only for log-det case $p = 0$:

Theorem (asympotic minimization[K. '21])

Let $r^* = \min_{X \in \mathcal{L}^{-1}(y)} \mathrm{rank}(X)$. *Then*

$$\mathcal{X}^* \subseteq \operatorname*{argmin}_{X^* \in \mathcal{L}^{-1}(y)} \prod_{j=1}^{r^*} \sigma_j(X^*) \quad \underline{\textit{subject to}} \quad \mathrm{rank}(X^*) = r^*$$

Search limits X^* of (global) minimizers X_γ for $\gamma \searrow 0$:

$$\mathcal{X}^* := \left\{ X^* \mid \exists (X_\gamma)_{\gamma>0}, \quad X^* = \lim_{\gamma \searrow 0} X_\gamma, \ X_\gamma \in \operatorname*{argmin}_{X \in \mathcal{L}^{-1}(y)} f_\gamma(X) \right\}$$

Only for log-det case $p = 0$:

Theorem (asympotic minimization[K. '21])

Let $r^* = \min_{X \in \mathcal{L}^{-1}(y)} \operatorname{rank}(X)$. *Then*

$$\mathcal{X}^* \subseteq \operatorname*{argmin}_{X^* \in \mathcal{L}^{-1}(y)} \prod_{j=1}^{r^*} \sigma_j(X^*) \quad \underline{\textit{subject to}} \quad \operatorname{rank}(X^*) = r^*$$

Process *solves* affine rank minimization (and IRLS-0 can minimize f_γ), but:

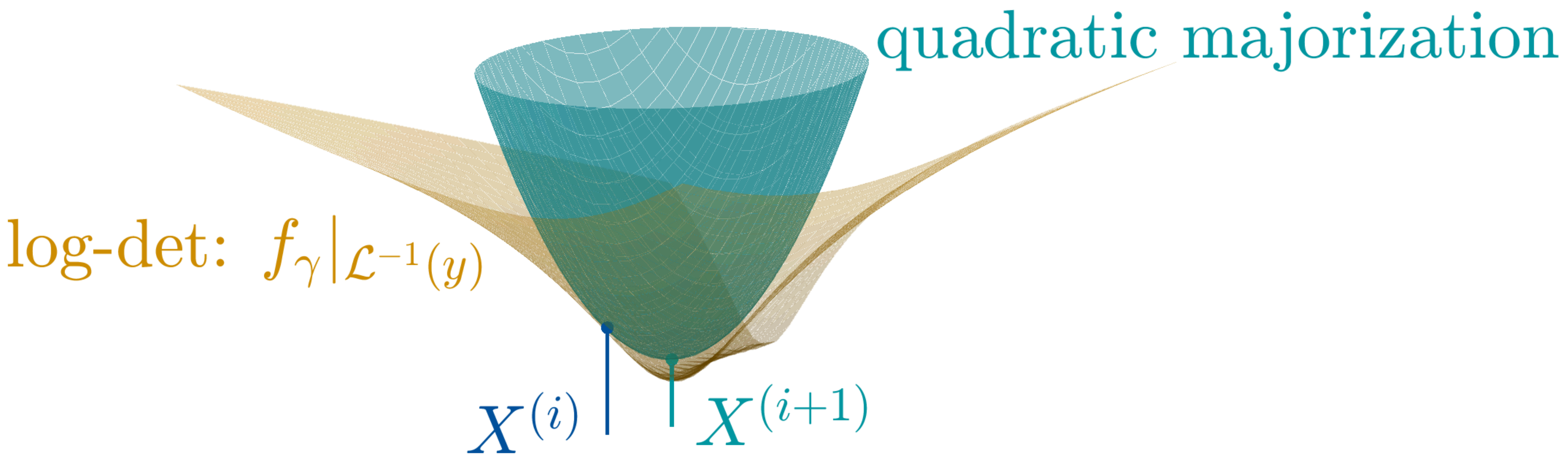
- **Q1:** prohibitiv local minima?
- **Q2:** numerically realize $\gamma \searrow 0$?

Classical, iterative quadratic majorization to minimize log-det:

Algorithm (classical matrix IRLS-0)

$$X^{(i+1)} \leftarrow \underset{X \in \mathcal{L}^{-1}(y)}{\operatorname{argmin}} \| W^{1/2}_{\gamma^{(i)}, X^{(i)}} X \|_F^2, \quad \gamma^{(i+1)} \leq \gamma^{(i)}, \quad i \in \mathbb{N}_0$$

for weights $W_{\gamma,X} := (XX^T + \gamma I_n)^{-1}$.

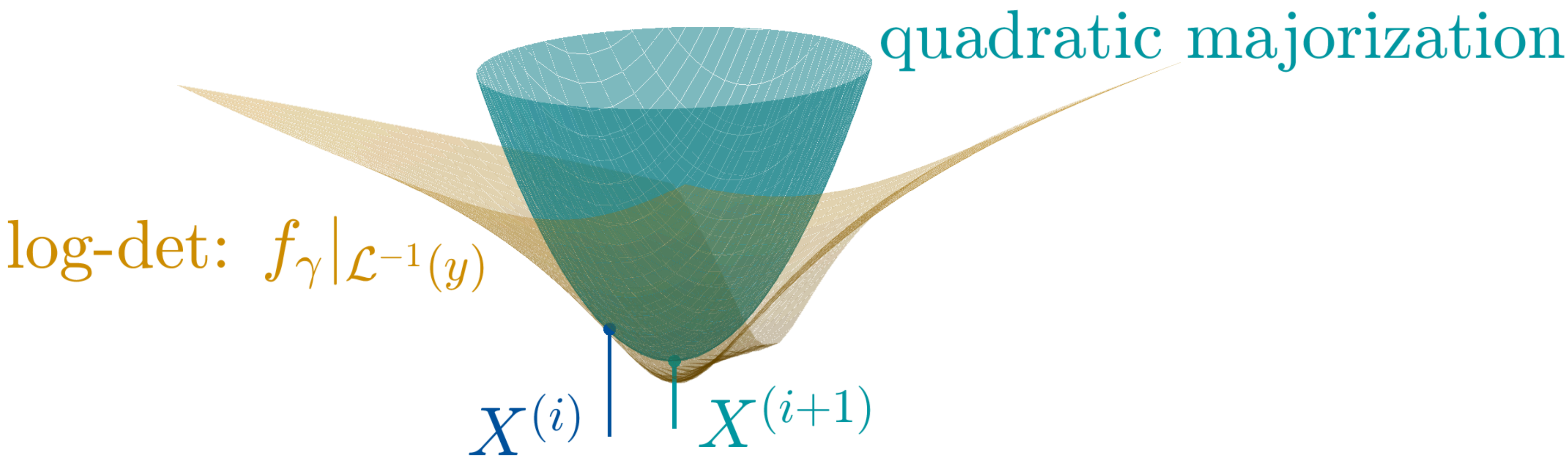


Classical, iterative quadratic majorization to minimize log-det:

Algorithm (classical matrix IRLS-0)

$$X^{(i+1)} \leftarrow \underset{X \in \mathcal{L}^{-1}(y)}{\operatorname{argmin}} \| W^{1/2}_{\gamma^{(i)}, X^{(i)}} X \|_F^2, \quad \gamma^{(i+1)} \le \gamma^{(i)}, \quad i \in \mathbb{N}_0$$

for weights $W_{\gamma,X} := (XX^T + \gamma I_n)^{-1}$.

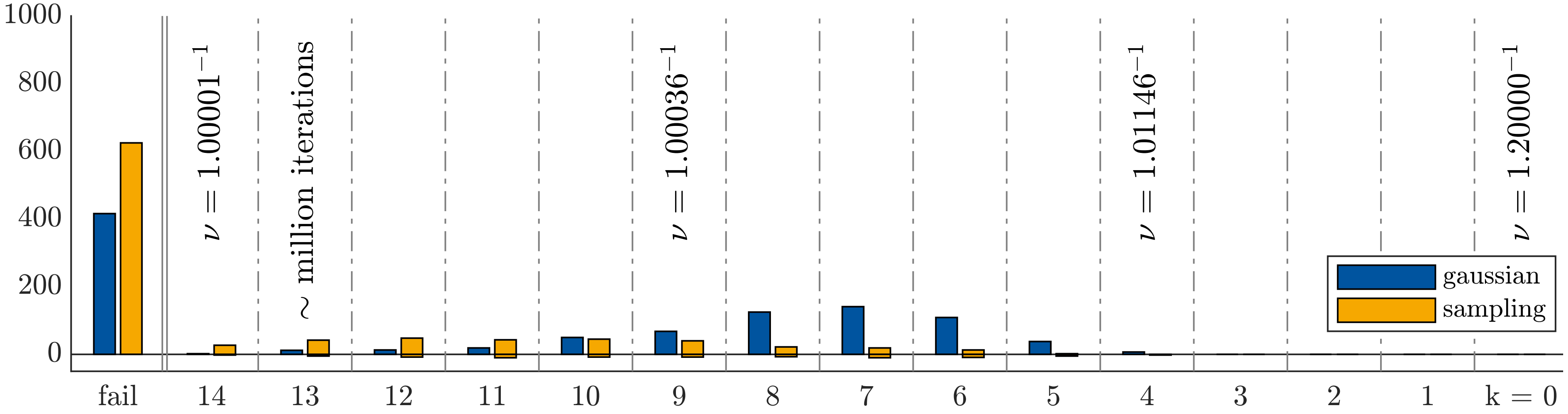


Experiment (classical matrix $\mathrm{IRLS\text{-}0}$ / 1000 instances)

- $y := \mathcal{L}(X^{(\mathrm{rs})}) \in \mathbb{R}^\ell$ via generic $\mathcal{L}$ and $X^{(\mathrm{rs})} : \mathrm{rank}(X^{(\mathrm{rs})}) = r$
- set $\gamma^{(i+1)} \leftarrow \nu\gamma^{(i)}$ for fixed $\nu < 1$ (but trying different $\nu = 1.2^{-1/2^k}$)
- $n = m = 12$, $r = 3$ for $\ell = \dim(V_{\leq r}) + 1 = 64$ measurements

Experiment (classical matrix IRLS-0 / 1000 instances)

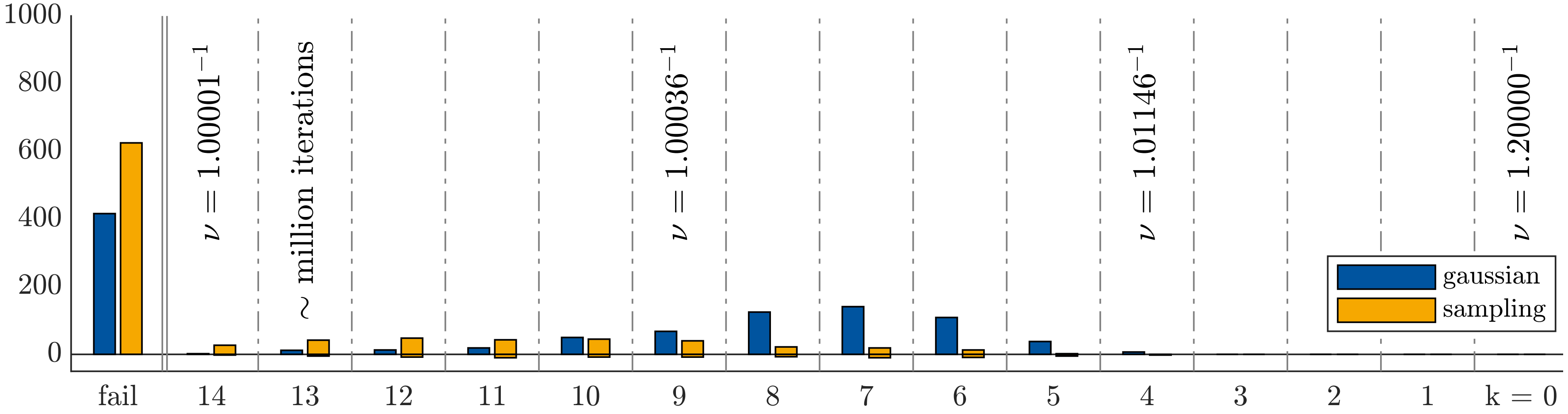
- $y := \mathcal{L}(X^{(\mathrm{rs})}) \in \mathbb{R}^\ell$ *via generic* $\mathcal{L}$ *and* $X^{(\mathrm{rs})} : \mathrm{rank}(X^{(\mathrm{rs})}) = r$
- *set* $\gamma^{(i+1)} \leftarrow \nu\gamma^{(i)}$ *for fixed* $\nu < 1$ *(but trying different* $\nu = 1.2^{-1/2^k}$*)*
- $n = m = 12$, $r = 3$ *for* $\ell = \dim(V_{\leq r}) + 1 = 64$ *measurements*



Affine rank minimization — Sensitivity experiment

Experiment (classical matrix IRLS-0 / 1000 instances)

- $y := \mathcal{L}(X^{(\mathrm{rs})}) \in \mathbb{R}^\ell$ via generic $\mathcal{L}$ and $X^{(\mathrm{rs})} : \mathrm{rank}(X^{(\mathrm{rs})}) = r$
- set $\gamma^{(i+1)} \leftarrow \nu\gamma^{(i)}$ for fixed $\nu < 1$ (but trying different $\nu = 1.2^{-1/2^k}$)
- $n = m = 12$, $r = 3$ for $\ell = \dim(V_{\leq r}) + 1 = 64$ measurements



- **@Q1:** local minima empirically less problematic
- **@Q2:** large potential in slow decrease $\gamma \searrow 0$ / better quadratic majorization

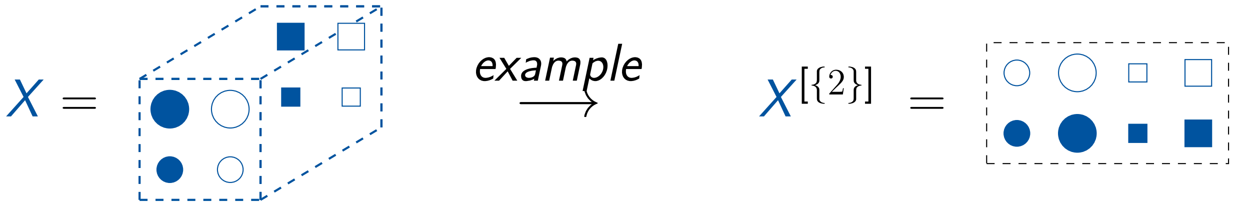
From Matrices to Tensors

Given:

- (surjective) linear operator $\mathcal{L} : \mathbb{R}^{n_1 \times \ldots \times n_d} \to \mathbb{R}^{\ell}$
- measurements $y \in \mathbb{R}^{\ell}$

Choose:

- family $\mathcal{K}$ of $S \subsetneq \{1, \ldots, d\}$ of matricizations



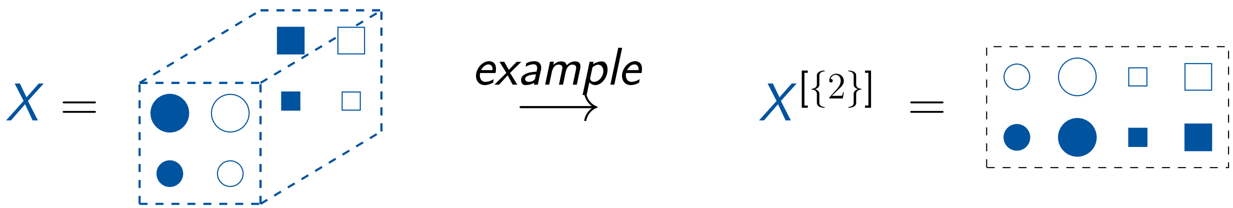
From Matrices to Tensors

Given:

- (surjective) linear operator $\mathcal{L} : \mathbb{R}^{n_1 \times \dots \times n_d} \to \mathbb{R}^{\ell}$
- measurements $y \in \mathbb{R}^{\ell}$

Choose:

- family $\mathcal{K}$ of $S \subsetneq \{1, \dots, d\}$ of matricizations



Definition (affine $\mathcal{K}$-sum-of-ranks minimization problem)

find $\quad X^* \in \underset{X \in \mathcal{L}^{-1}(y)}{\operatorname{argmin}} \sum_{S \in \mathcal{K}} \operatorname{rank}(X^{[S]}), \qquad f_\gamma^{\mathcal{K}}(X) := \sum_{S \in \mathcal{K}} f_\gamma(X^{[S]})$

Affine sum-of-ranks minimization — Phase transition

Experiment (tensor $\mathrm{IRLS}\text{-}0\mathcal{K}$)

- $\mathcal{K} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}\}$
- $y := \mathcal{L}(X^{(\mathrm{rs})})$ *for generic* $\mathcal{L}$ *and* $X^{(\mathrm{rs})} : \mathrm{rank}(X^{(\mathrm{rs})[S]}) = \bar{r}$, $S \in \mathcal{K}$
- $d = 4$, $\bar{n} = 5$, $\bar{r} = 3 \to \dim(\mathcal{V}^{\mathcal{K}}_{\leq \bar{r}}) = 69$
- *suitable realization of* $\gamma \searrow 0$

Affine sum-of-ranks minimization — Phase transition

Experiment (tensor IRLS-0$\mathcal{K}$)

- $\mathcal{K} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}\}$
- $y := \mathcal{L}(X^{(\mathrm{rs})})$ *for generic* $\mathcal{L}$ *and* $X^{(\mathrm{rs})}$: $\mathrm{rank}(X^{(\mathrm{rs})[S]}) = \overline{r}$, $S \in \mathcal{K}$
- $d = 4$, $\overline{n} = 5$, $\overline{r} = 3 \to \dim(\mathcal{V}^{\mathcal{K}}_{\leq \overline{r}}) = 69$
- *suitable realization of* $\gamma \searrow 0$

instance:		recovery rate in %
$\ell = 90$	gaussian	100
	sampling	78
$\ell = 80$	gaussian	88
	sampling	43
$\ell = 70$	gaussian	18
	sampling	7

Affine sum-of-ranks minimization — Phase transition

Experiment (tensor $\mathrm{IRLS\text{-}0}\mathcal{K}$)

- $\mathcal{K} = \{\{1\},\{2\},\{3\},\{4\},\{1,2\}\}$
- $y := \mathcal{L}(X^{(\mathrm{rs})})$ *for generic* $\mathcal{L}$ *and* $X^{(\mathrm{rs})} : \mathrm{rank}(X^{(\mathrm{rs})[S]}) = \overline{r},\ S \in \mathcal{K}$
- $d = 4,\ \overline{n} = 5,\ \overline{r} = 3 \to \dim(\mathcal{V}^{\mathcal{K}}_{\leq \overline{r}}) = 69$
- *suitable realization of* $\gamma \searrow 0$

instance:		recovery rate in %
$\ell = 90$	● gaussian	100
	● sampling	78
$\ell = 80$	● gaussian	88
	● sampling	43
$\ell = 70$	● gaussian	18
	● sampling	7

For ● 100%, sum-of-nuclear-norms ($p = 1$) needs $\ell \geq 400$ (of $d^{\overline{n}} = 625$ entries).

Alternating optimization on tensor tree networks — AIRLS-$0\mathcal{K}$ requirements

From Small Tensors to High Dimensions

- relaxation of affine constraint:

$$F_\gamma^{\mathcal{K}}(X) := \|\mathcal{L}(X) - y\|_F^2 + \gamma \cdot a_\gamma \sum_{S \in \mathcal{K}} f_\gamma(X^{[S]})$$

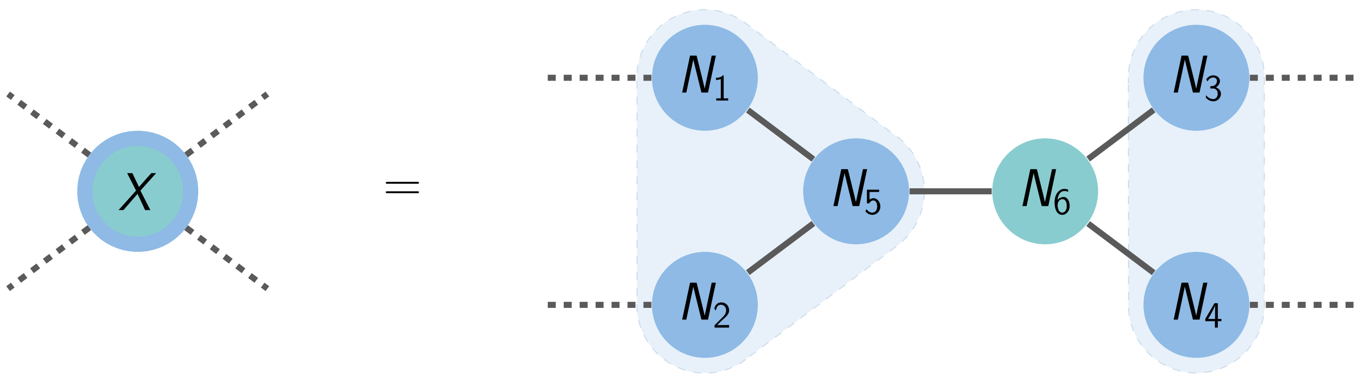
Alternating optimization on tensor tree networks — AIRLS-0$\mathcal{K}$ requirements

From Small Tensors to High Dimensions

- relaxation of affine constraint:

$$F_{\gamma}^{\mathcal{K}}(X) := \|\mathcal{L}(X) - y\|_F^2 + \gamma \cdot a_{\gamma} \sum_{S \in \mathcal{K}} f_{\gamma}(X^{[S]})$$

- assume $\mathcal{K}$ hierarchical (TT/MPS, Tucker/HOSVD, HT) and represent $X = \tau(N)$

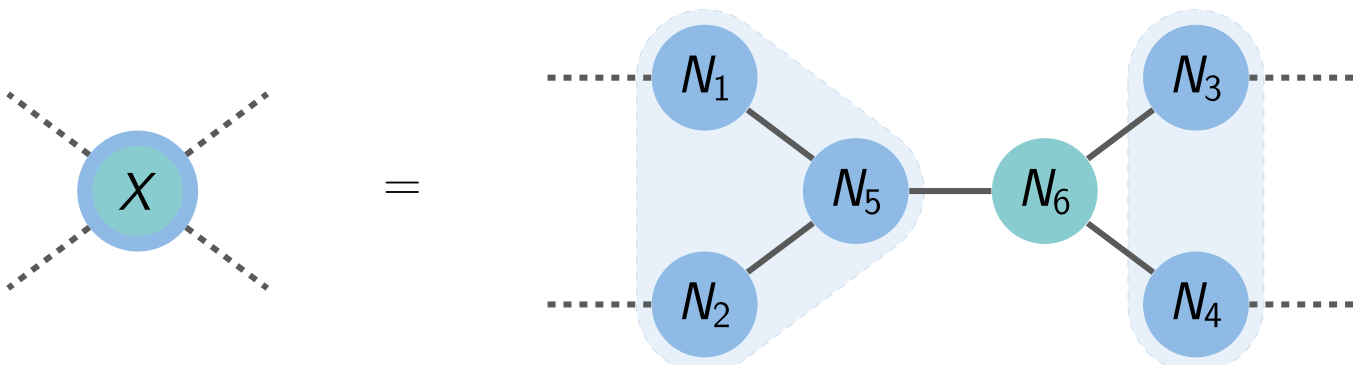


From Small Tensors to High Dimensions

- relaxation of affine constraint:

$$F_\gamma^{\mathcal{K}}(X) := \|\mathcal{L}(X) - y\|_F^2 + \gamma \cdot a_\gamma \sum_{S \in \mathcal{K}} f_\gamma(X^{[S]})$$

- assume $\mathcal{K}$ hierarchical (TT/MPS, Tucker/HOSVD, HT) and represent $X = \tau(N)$



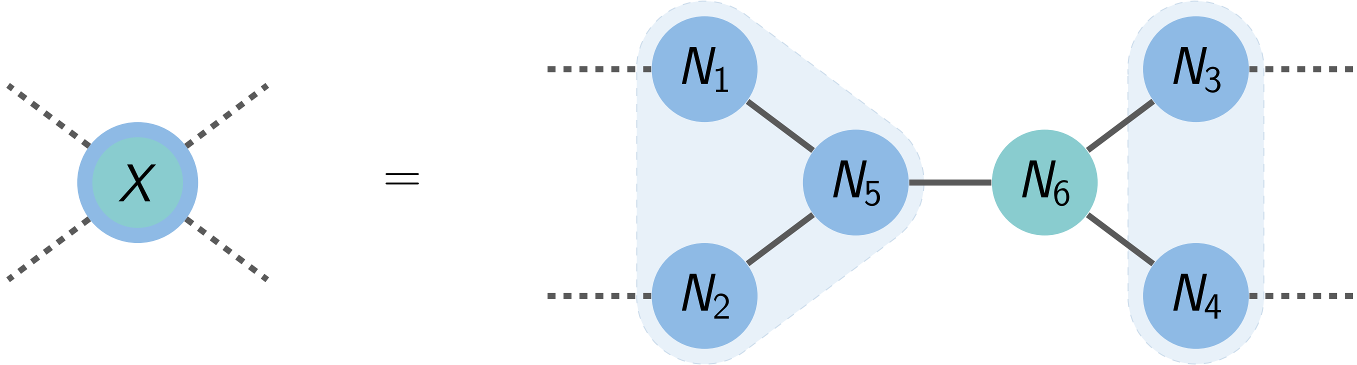
- operator $\mathcal{L}$ with low $\mathcal{K}$-ranks (e.g. sampling has rank 1)

From Small Tensors to High Dimensions

- relaxation of affine constraint:

$$F_\gamma^{\mathcal{K}}(X) := \|\mathcal{L}(X) - y\|_F^2 + \gamma \cdot a_\gamma \sum_{S \in \mathcal{K}} f_\gamma(X^{[S]})$$

- assume $\mathcal{K}$ hierarchical (TT/MPS, Tucker/HOSVD, HT) and represent $X = \tau(N)$



- operator $\mathcal{L}$ with low $\mathcal{K}$-ranks (e.g. sampling has rank 1)
- alternatingly minimize quadratic majorizations of $F_\gamma^{\mathcal{K}}(X)$ via $X = \tau(N)$
  $\rightarrow$ possible in same, **polynomial complexity**[K. '21] as evaluating $\mathcal{L}(\tau(N))$

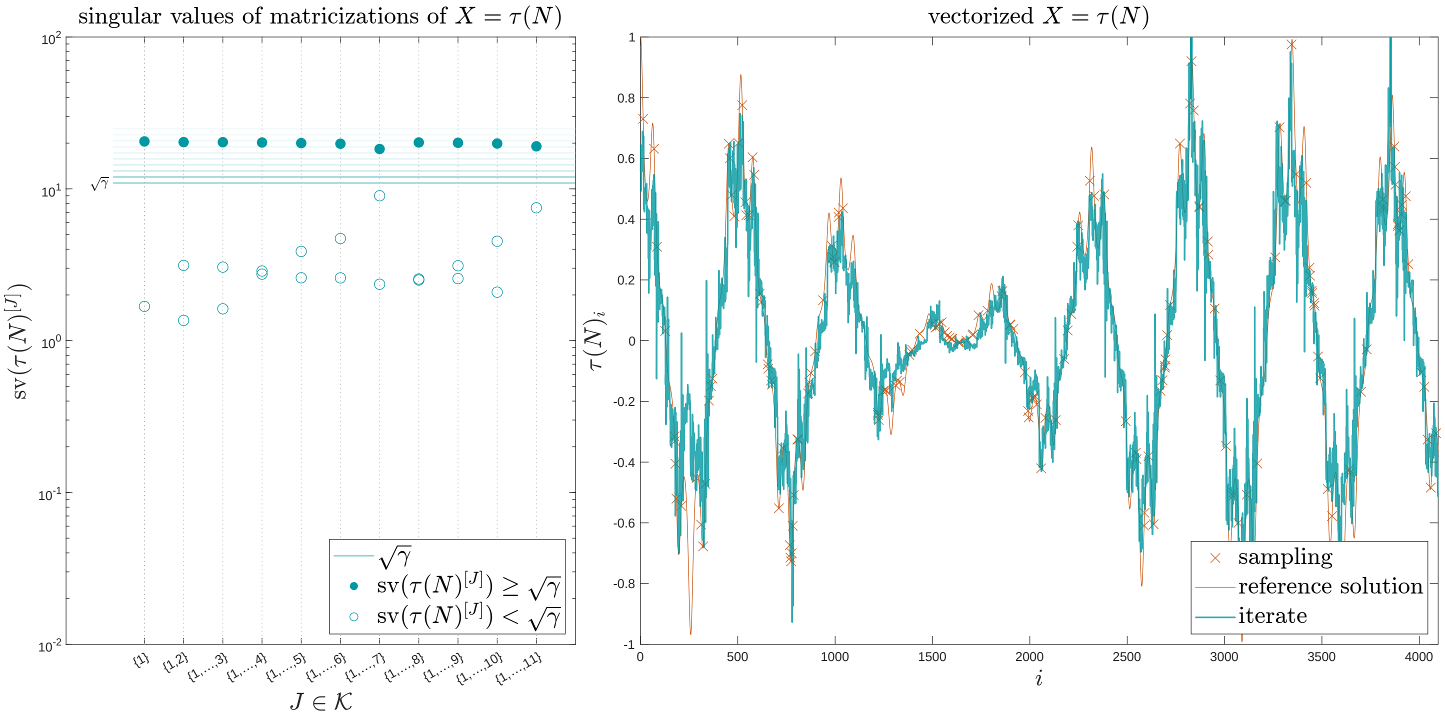
Alternating optimization on tensor tree networks — Exemplary application

Experiment (QTT-interpolation with $\mathrm{AIRLS}\text{-}0\mathcal{K}$)

- *initial teaser: approximate function given at* 180 *values (with* 1‰ *white noise)*
- *tensorization from* $g \in \mathbb{R}^{2^{12}}$ *to* $X \in \mathbb{R}^{2\times\ldots\times 2}$ *turns interpolation into* $\mathcal{L}(X) = y$
- *tensor train:* $\mathcal{K} := \{\{1\},\ \{1,2\},\ \ldots,\ \{1,\ldots,11\}\}$

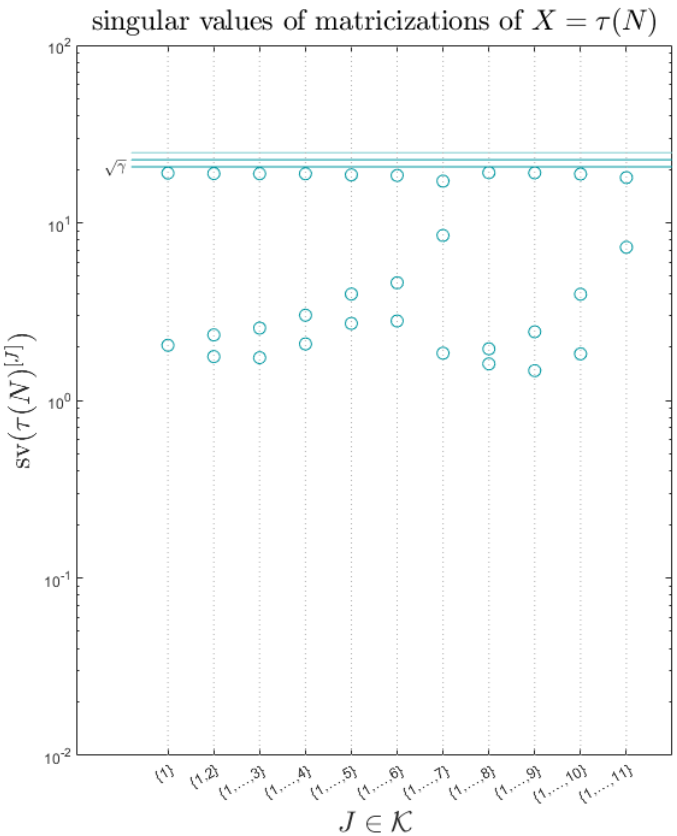
Experiment (QTT-interpolation with $\mathrm{AIRLS}\text{-}0\mathcal{K}$)

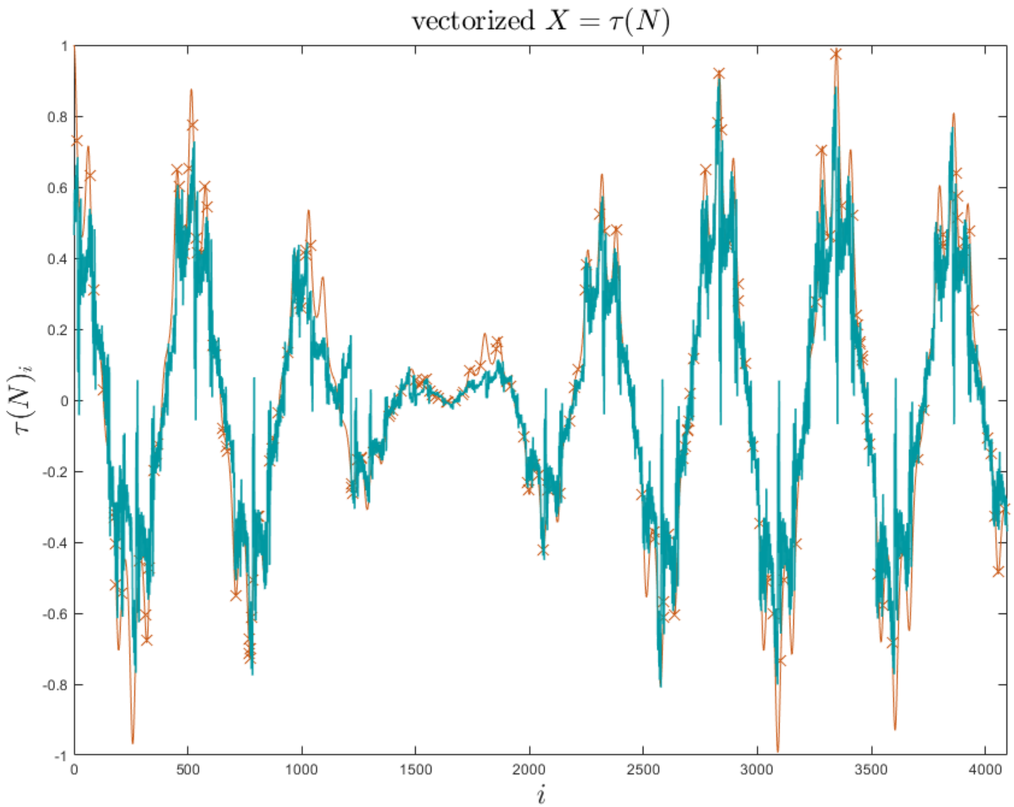
- *initial teaser: approximate function given at 180 values (with 1‰ white noise)*
- *tensorization from* $g \in \mathbb{R}^{2^{12}}$ *to* $X \in \mathbb{R}^{2\times\ldots\times 2}$ *turns interpolation into* $\mathcal{L}(X) = y$
- *tensor train:* $\mathcal{K} := \{\{1\},\ \{1,2\},\ \ldots,\ \{1,\ldots,11\}\}$



Experiment (QTT-interpolation with $\mathrm{AIRLS}\text{-}0\mathcal{K}$)

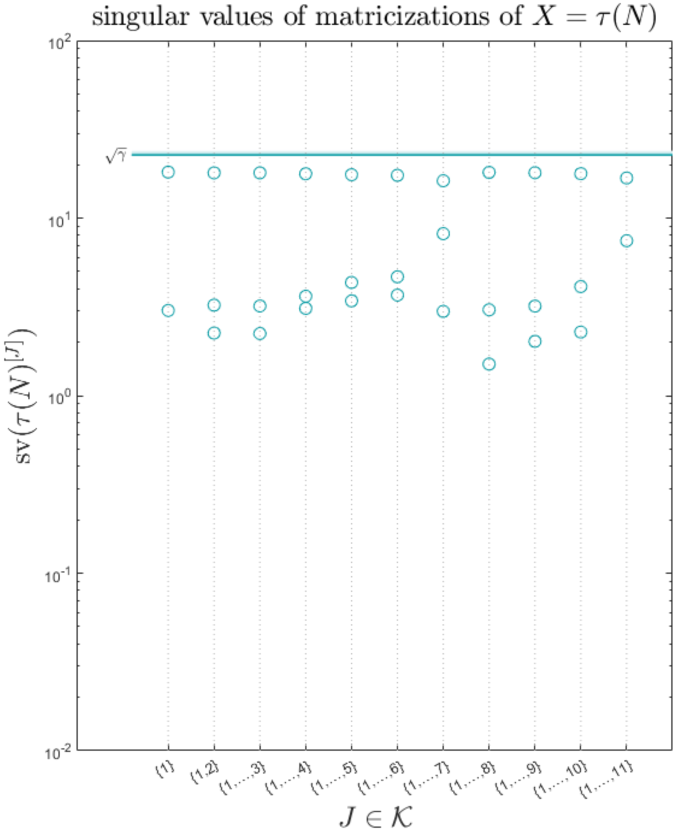
- *first: too fast decrease of* γ

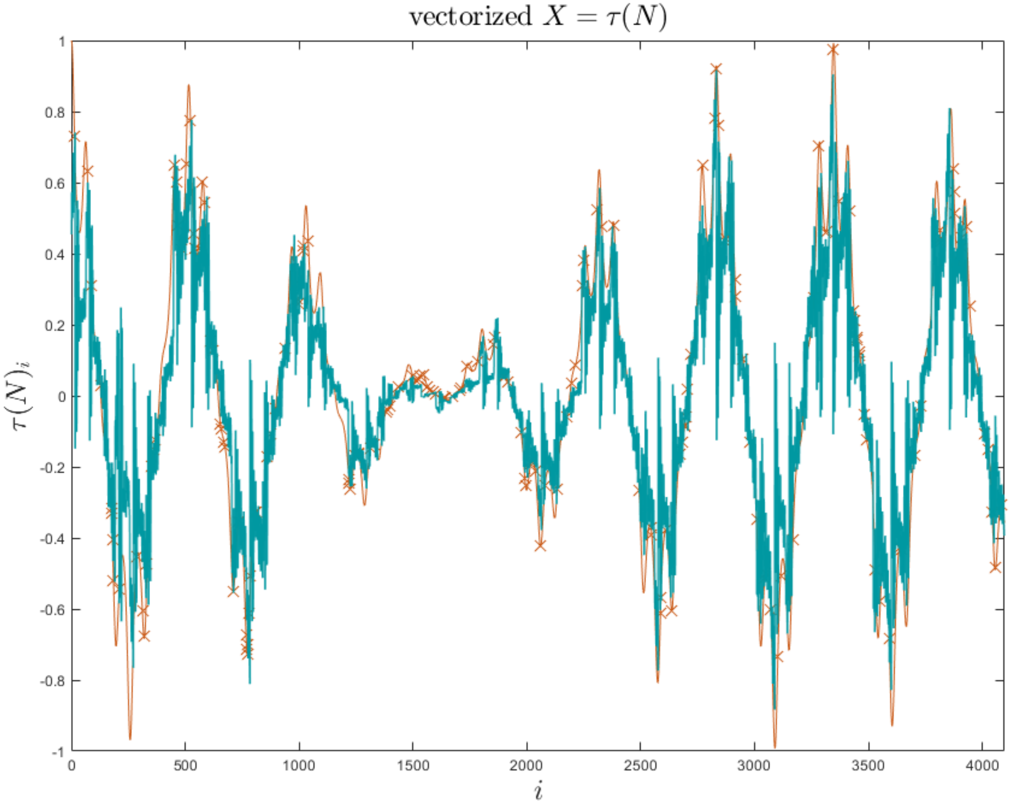




Experiment (QTT-interpolation with AIRLS-0$\mathcal{K}$)

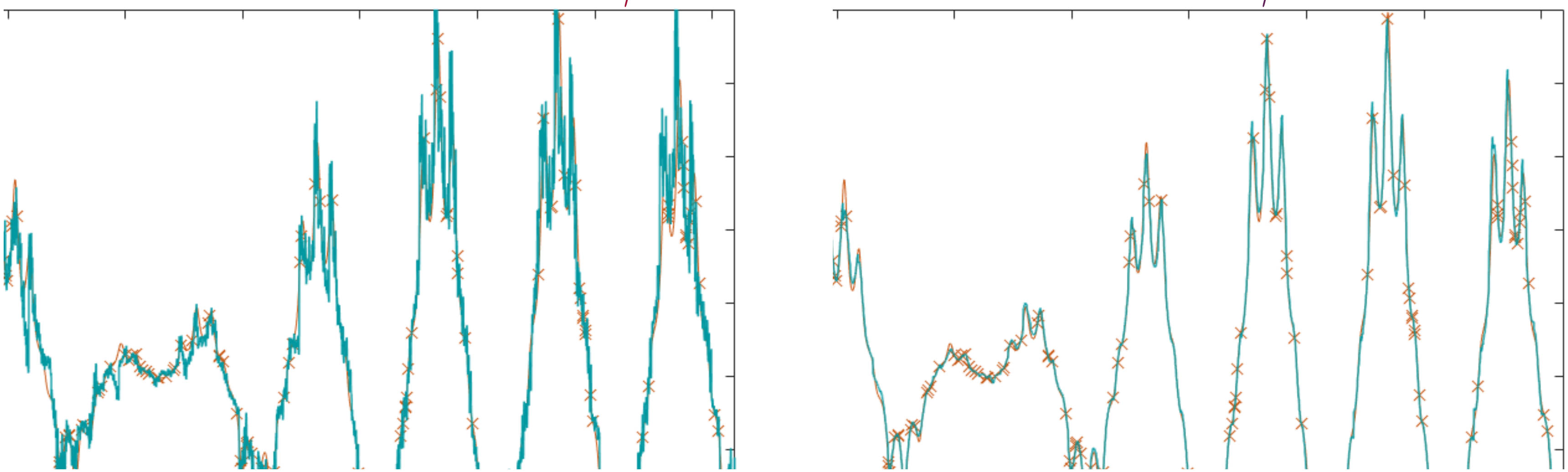
- *slower decrease of* γ





Conclusions — Comparison and Future directions

fast decrease of γ | slow decrease of γ



(Future directions:)

- improve on theoretical bounds (in particular exclusion of local minima)
- derive and prove *optimal quadratic majorization*[cf. Kümmerle and Verdun '21]
- generalize to low-rank interpolation of multivariate functions[K. '20]
- adapt (as heuristic) to cyclic networks (e.g. cyclic MPS, PEPS)