Iterative hard thresholding based algorithms for low-rank tensor recovery

2 authors:

José Henrique de Morais Goulart
GIPSA-lab
19 PUBLICATIONS **52** CITATIONS

SEE PROFILE

Gerard Favier
University of Nice Sophia Antipolis
315 PUBLICATIONS **2,059** CITATIONS

SEE PROFILE

Iterative hard thresholding based algorithms for low-rank tensor recovery

José Henrique de Morais Goulart
In collaboration with : Gérard Favier

I3S Laboratory, Sophia Antipolis, France

i3S sophia antipolis

Université Nice Sophia Antipolis

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico

Tensor-related research @ I3S

- Tensor models (de Almeida, da Costa, Favier)
  - Block constrained CPD
  - Generalized Paratuck
  - Nested CPD
  - Nested Tucker
  - Overview of constrained CPD models
  - 7 j. papers, 1 book chapter
- Estimation of structured CPD models (Goulart, Cohen, Boyer, Boizard, Favier, Kibangou, Comon)
  - 4 conf. papers, 2 j. papers
- Tensor completion (Goulart, Favier):
  - 1 conf. paper, 1 j. paper (submitted)

Tensor-related research @ I3S

- System identication (Kibangou, Khouaja, Fernandes, Bouilloc, Favier)
  - HOS-based linear system identication
  - Nonlinear system modeling and identication : Block structured systems (Wiener, Hammerstein, W-H), Volterra systems
  - 13 j. papers
- SAR image processing (Porges, Thales)
  - 2 conf. papers
- Wireless communications:
  - MIMO nonlinear systems (A. Fernandes, Favier)
  - MIMO point-to-point systems (de Almeida, Bouilloc, da Costa, Favier)
  - MIMO cooperative relay systems (Ximenes, de Almeida, Freitas, Favier)
  - 4 book ch., >20 j. papers, >30 conf. papers
- Tensor completion for traffic data estimation (Goulart, Kibangou, Favier)

Low-rank tensor recovery (LRTR)

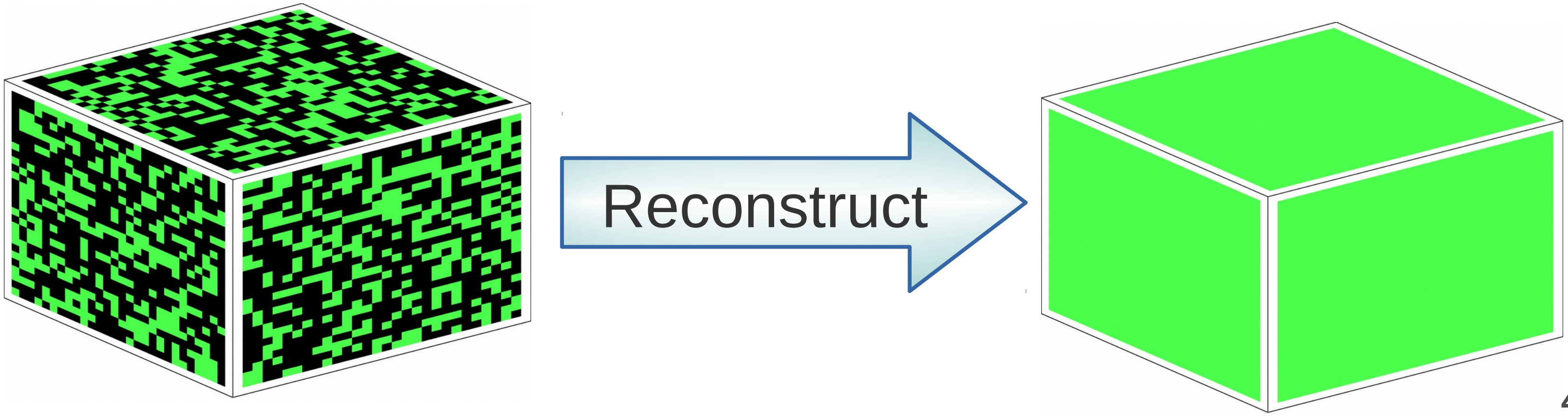
- Recover $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N_1 \times \cdots \times N_P}$ from

$$\mathbf{y} = \mathscr{A}(\boldsymbol{\mathcal{X}}) + \mathbf{e} \in \mathbb{R}^{M}, \quad M < N_1 \ldots N_P$$

$\mathscr{A}$ → linear measurement operator (MO)

Premise : $\boldsymbol{\mathcal{X}}$ has low rank

- Most usual setting : Tensor Completion (TC)



Which rank?

- Tensor rank : CPD model

$$\mathcal{X} = \sum_{r}^{R} \bigotimes_{p=1}^{P} \mathbf{a}_r^{(p)} \longrightarrow \boxed{\text{DOF} \quad \mathcal{O}(PNR)}$$

- Multilinear rank : Tucker model

$$\mathcal{X} = \underbrace{\mathcal{G}}_{R_1 \times R_2 \times \cdots \times R_P} \times_{p=1}^{P} \mathbf{U}^{(p)} \longrightarrow \boxed{\text{DOF} \quad \mathcal{O}(R^P + PNR)}$$

- Ideally : recovery from $M = \mathcal{O}(\mathrm{DOF})$

- Typical sampling bounds : $\mathcal{O}(RN^{P-1})$

Main approaches

convex

- Minimizing sum of nuclear norms (SNN)

$$\min_{\mathcal{X}} \|\mathbf{y} - \mathscr{A}(\mathcal{X})\|_2^2 + \lambda \sum_{p=1}^{P} \gamma_p \left\|\mathbf{X}_{\langle p \rangle}\right\|_*$$

- Tensor nuclear norm : conditional gradient

$$\min_{\mathcal{X}} J(\mathcal{X}) + \lambda \|\mathcal{X}\|_*$$

  - search direction : best rank-one approx.

non-convex

- Low-rank matrix factorization of unfoldings

$$\min_{\mathcal{X}, \mathbf{W}_p, \mathbf{Z}_p} \sum_p^P \alpha_p \left\| \underbrace{\mathbf{W}_p}_{N_p \times R_p} \underbrace{\mathbf{Z}_p}_{R_p \times \bar{N}_p} - \mathbf{X}_{\langle p \rangle} \right\|_F^2 \text{ s.t. } \mathscr{A}(\mathcal{X}) = \mathbf{y}$$

- Constrained least-squares : $\min_{\mathcal{X} \in \mathcal{S}} J(\mathcal{X})$
  - Riemannian opt., iterative hard thresholding

Iterative hard thresholding (IHT)

$$\min_{\mathcal{X} \in \mathcal{L}_{\mathbf{r}}} \frac{1}{2} J(\mathcal{X})$$

$$\nabla_J(\mathcal{X}) = -\mathscr{A}^* \left(\mathbf{y} - \mathscr{A}\left(\mathcal{X} \right) \right)$$

$$\mathcal{X}_{k+1} = \mathscr{H}_{\mathbf{r}} \left(\mathcal{X}_k - \mu_k \nabla_J(\mathcal{X}_k) \right)$$

$$\mathcal{L}_{\mathbf{r}} = \mathcal{L}_{(R_1,\ldots,R_P)} = \{ \mathcal{X} : \operatorname{rank}(\mathbf{X}_{\langle p \rangle}) \leq R_p \}$$

Ideally : HT operator projects onto $\mathcal{L}_{\mathbf{r}}$

Intractable ⇒ approximate projection

- Desirable properties :
  1) Accuracy (*e.g.*, bounded error)
  2) Low computing cost
  3) Analytical tractability

Tensor IHT (TIHT) [Rauhut 2013]

- HT : truncated HOSVD [De Lathauwer 2000]
  - Projection onto dominant modal subspaces

$$\mathscr{H}_{\mathbf{r}}(\boldsymbol{\mathfrak{X}}) = \boldsymbol{\mathfrak{X}} \times_{p=1}^{P} \mathbf{U}^{(p)} (\mathbf{U}^{(p)})^T, \qquad \mathbf{U}^{(p)} \in \mathbb{R}^{N_p \times R_p}$$

  - Quasi-optimal

$$\|\boldsymbol{\mathfrak{X}} - \mathscr{H}_{\mathbf{r}}(\boldsymbol{\mathfrak{X}})\|_F \leq \sqrt{P} \min_{\boldsymbol{\mathfrak{Z}} \in \mathcal{L}_{\mathbf{r}}} \|\boldsymbol{\mathfrak{X}} - \boldsymbol{\mathfrak{Z}}\|_F$$

  - Complexity

$$\mathcal{O}\left(\sum_p N_p \bar{N}_p \min\{N_p, \bar{N}_p\}\right) \qquad \bar{N}_p = \prod_{q \neq p} N_q$$

  - Suboptimality makes the analysis hard
    - Needs additional assumptions

SeMPIHT algorithm

$$\boxed{\mathcal{X}_{k+1} = \mathcal{S}_{\mathbf{r}} \left(\mathcal{X}_k + \mu_k \mathcal{A}^* \left(\mathbf{y} - \mathcal{A}(\mathcal{X}_k)\right)\right)}$$

- Sequentially optimal projections onto dominant subspaces [Vannieuwenhoven 2012]

$$\mathcal{X} \xrightarrow{\mathbf{\Pi}_{R_1}} \mathcal{V}_1 \xrightarrow{\mathbf{\Pi}_{R_2}} \cdots \xrightarrow{\mathbf{\Pi}_{R_{P-1}}} \mathcal{V}_{P-1} \xrightarrow{\mathbf{\Pi}_{R_P}} \mathcal{V}_P = \mathcal{S}_{\mathbf{r}}(\mathcal{X})$$

- Cheaper, as dimensions can be gradually reduced

$$\mathcal{V}_p = \mathcal{V}_{p-1} \times_p \left(\mathbf{U}^{(p)}\right)^T$$

$$\hookrightarrow R_1 \times \cdots \times R_p \times N_{p+1} \times \cdots \times N_P$$

- Quasi-optimal

$$\|\mathcal{X} - \mathcal{S}_{\mathbf{r}}(\mathcal{X})\|_F \leq \sqrt{P} \min_{\mathcal{Z} \in \mathcal{L}_{\mathbf{r}}} \|\mathcal{X} - \mathcal{Z}\|_F$$

SeMPIHT : step size choice

$$\boxed{\mathcal{X}_{k+1} = \mathcal{S}_{\mathbf{r}}\left(\mathcal{X}_k + \mu_k \mathcal{A}^*\left(\mathbf{y} - \mathcal{A}(\mathcal{X}_k)\right)\right)}$$

- Improved step size (ISS) heuristic [Goulart 2015]

$$\alpha\,\omega(\mathcal{X}_{k+1}, \mathcal{X}_k) \leq \mu_k < \omega(\mathcal{X}_{k+1}, \mathcal{X}_k), \quad \alpha < 1$$

TIHT [Rauhut 2013]

$$\mu_k = \frac{\|\nabla J\|_F^2}{\|\mathcal{A}(\nabla J)\|_2^2}$$

- Often, too small steps

NTIHT [Rauhut 2016]

$$\mu_k = \|\mathcal{G}_k\|_F^2 \left\|\mathcal{A}\left(\mathcal{G}_k\right)\right\|_2^{-2}$$

$$\mathcal{G}_k = \nabla J \times_{p=1}^{P} \mathbf{U}^{(p)} \mathbf{U}^{(p)T}$$

- Comparable to ISS

Analysis of SeMPIHT

- Exploits sequential optimality of modal proj.
- Based on Restricted Isometry Property (RIP)

$$\forall \boldsymbol{\mathcal{X}} \in \mathcal{L}_{\mathbf{r}}, \quad (1-\delta_{\mathbf{r}})\|\boldsymbol{\mathcal{X}}\|_F^2 \leq \|\mathscr{A}(\boldsymbol{\mathcal{X}})\|_2^2 \leq (1+\delta_{\mathbf{r}})\|\boldsymbol{\mathcal{X}}\|_F^2$$

Theorem [Goulart 2016 (submitted)] :

- If $\delta_{\mathbf{r}_{3,p}} < 2^{-P}, \qquad \mathbf{r}_{3,p} = (N_1, \ldots, 3R_p, \ldots, N_P),$

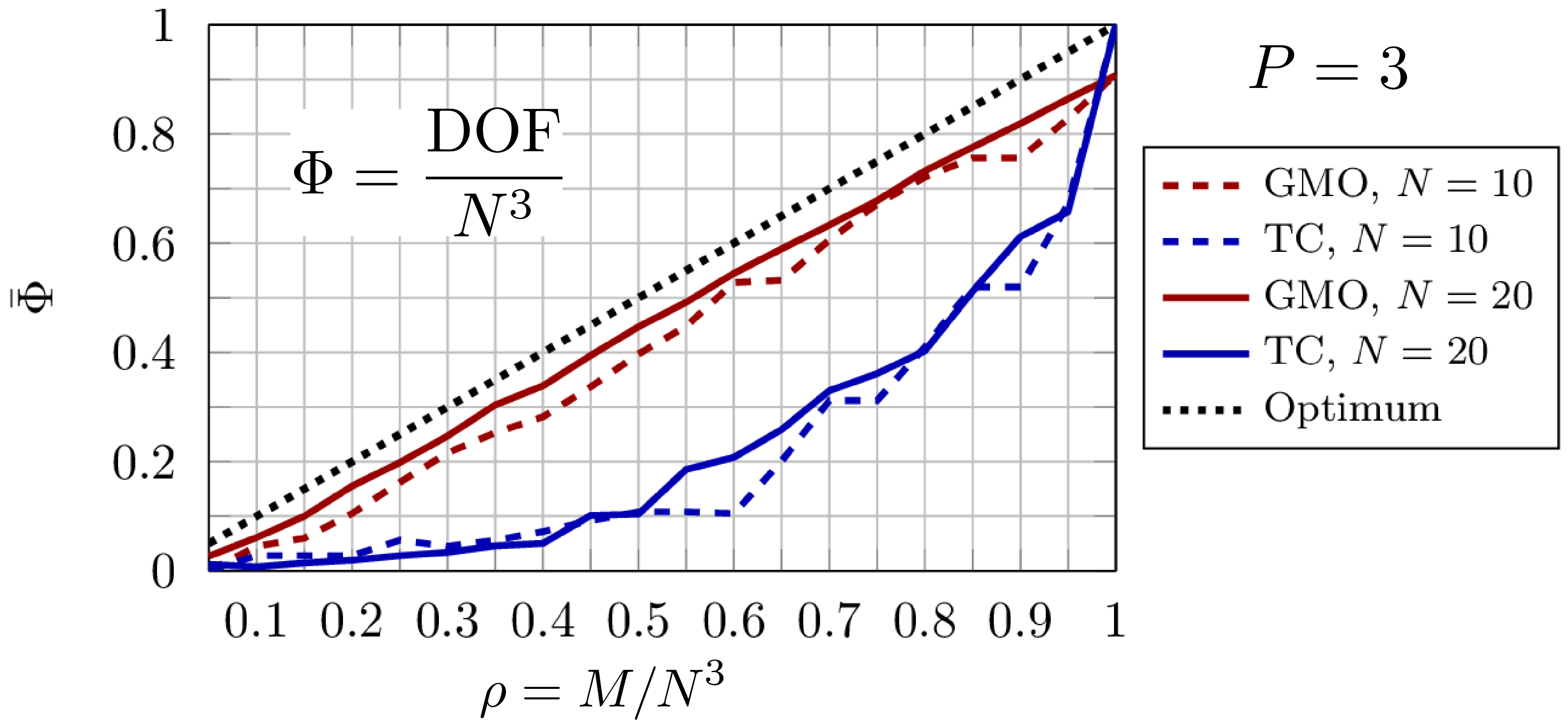
$$\|\boldsymbol{\mathcal{X}}_{\mathbf{r}} - \boldsymbol{\mathcal{X}}_k\|_F \leq \xi^k \|\boldsymbol{\mathcal{X}}_{\mathbf{r}} - \boldsymbol{\mathcal{X}}_0\|_F + \gamma \left\|\mathscr{A}(\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}_{\mathbf{r}}) + \mathbf{e}\right\|_2$$

$$\xi = 2^P \delta_{\mathbf{r}_{3,p}} < 1 \qquad \gamma = 2^P \sqrt{1-\delta_{\mathbf{r}_{3,p}}}/(1-\xi)$$

$$\boldsymbol{\mathcal{X}}_{\mathbf{r}} \in \arg\min_{\boldsymbol{\mathcal{Z}} \in \mathcal{L}_{\mathbf{r}}} \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Z}}\|_F$$

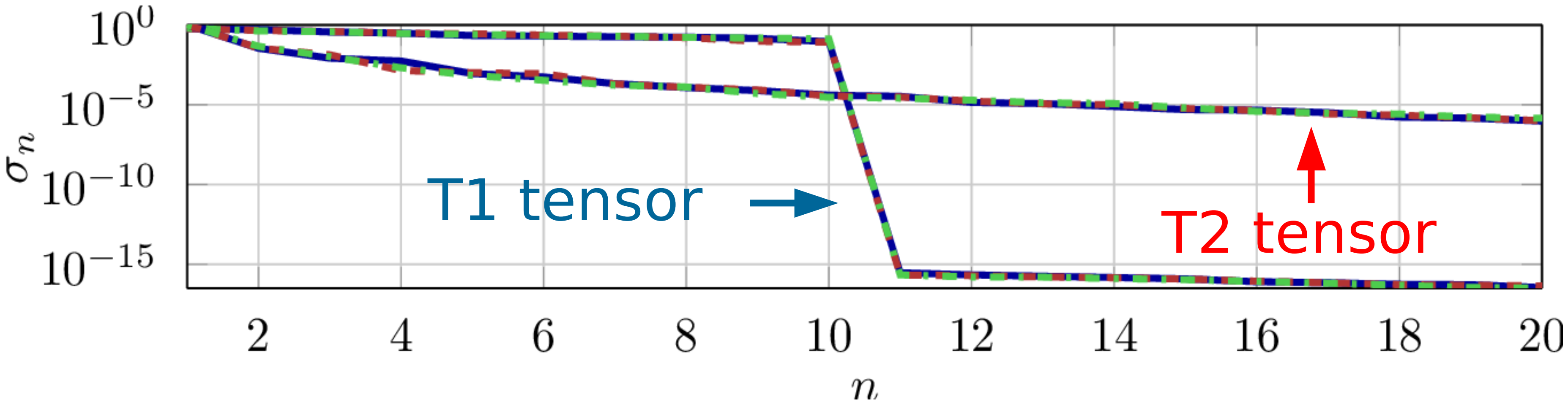
Sampling bounds

- Suboptimal : $\mathcal{O}(RN^{P-1})$
- Empirically : optimal for Gaussian MO



Experimental evaluation

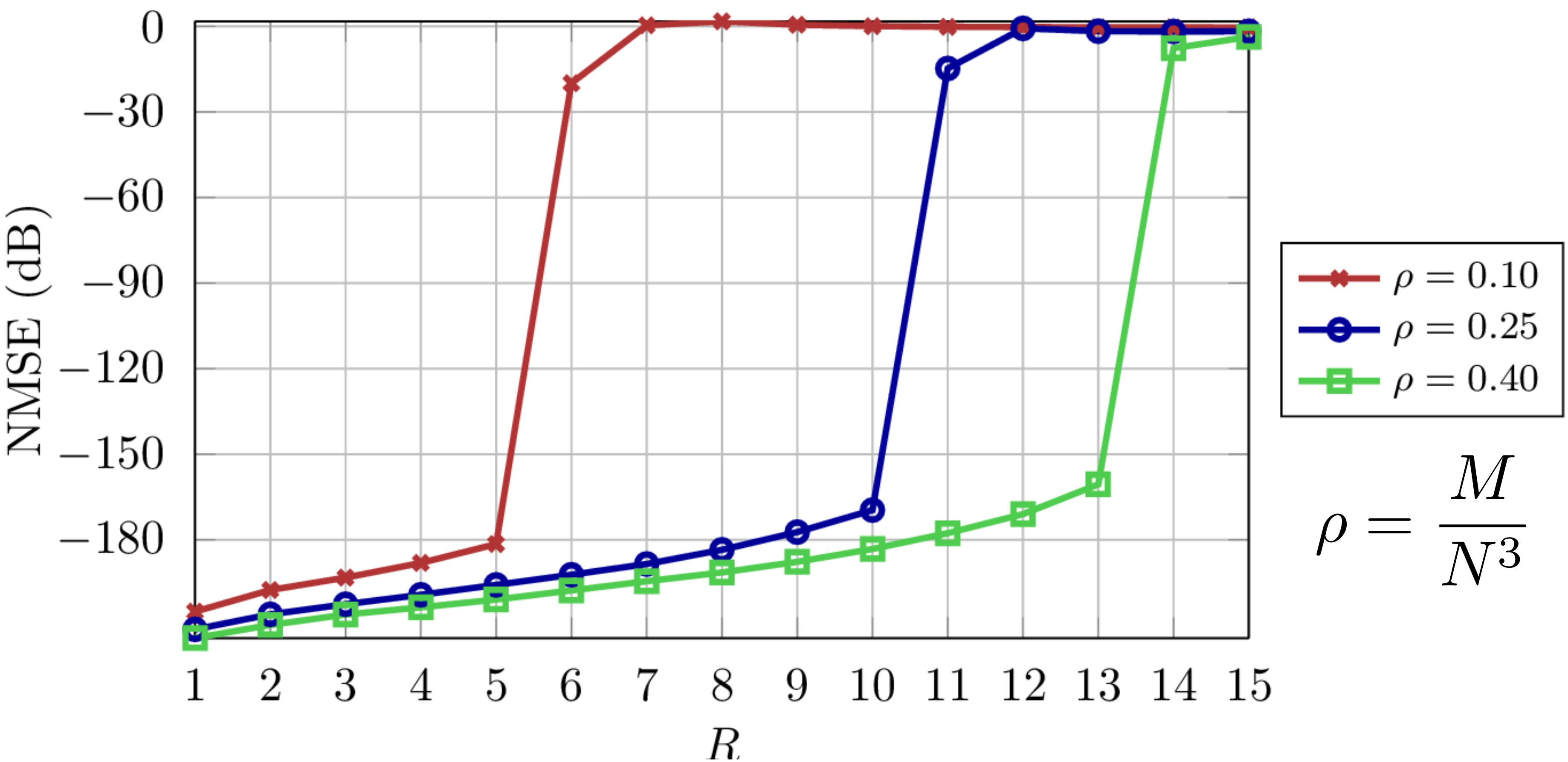
- Measurement operators :
  - (1) Gaussian MOs ; (2) Sampling MOs (TC)
- Random tensor classes :
  - T1 tensors : mrank exactly low
  - T2 tensors : fast decaying modal spectra



- Criterion : $\mathrm{NMSE} = \|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 / \|\mathcal{X}\|_F^2$

Recovery performance (Gauss. MO)

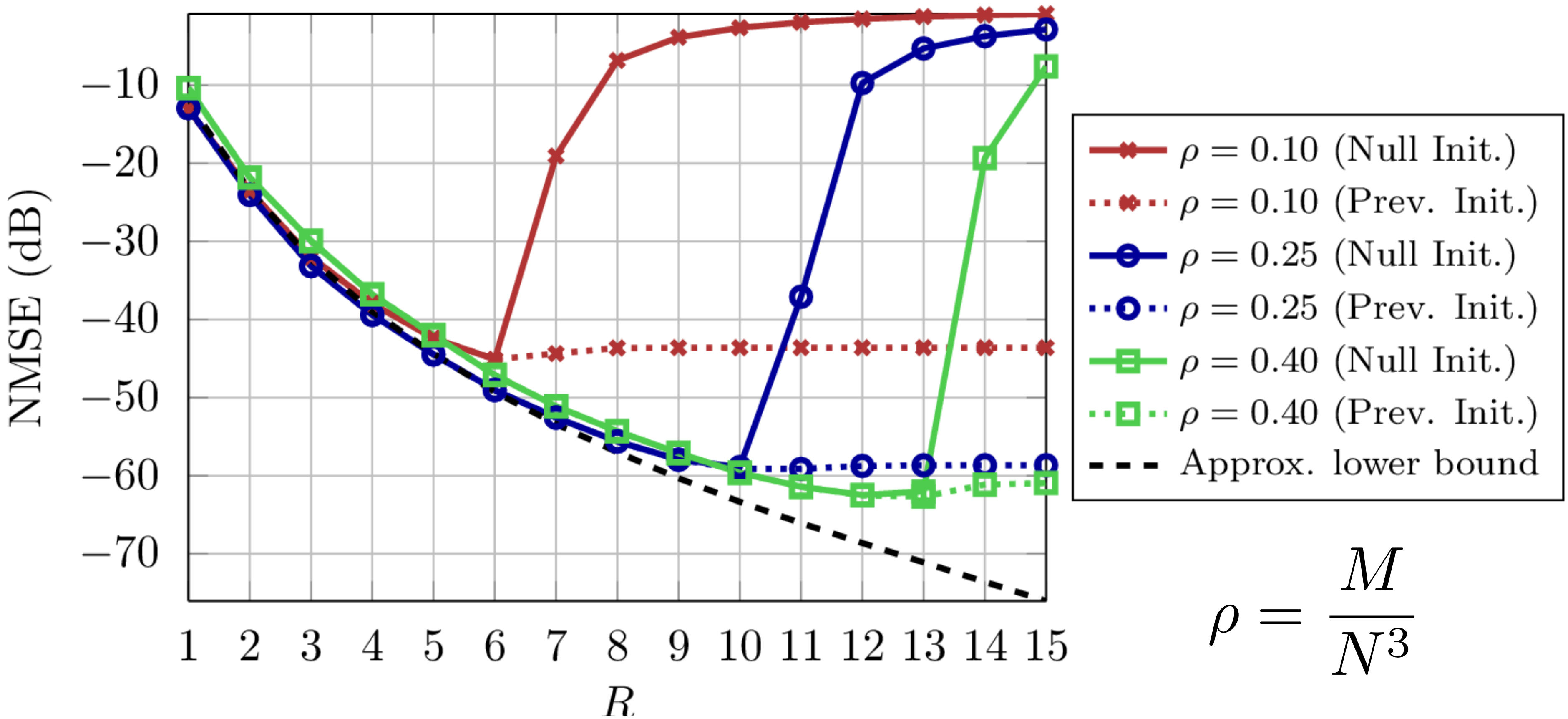
- Recovery of 20x20x20 T1 tensors



$$\rho = \frac{M}{N^3}$$

Recovery performance (Gauss. MO)

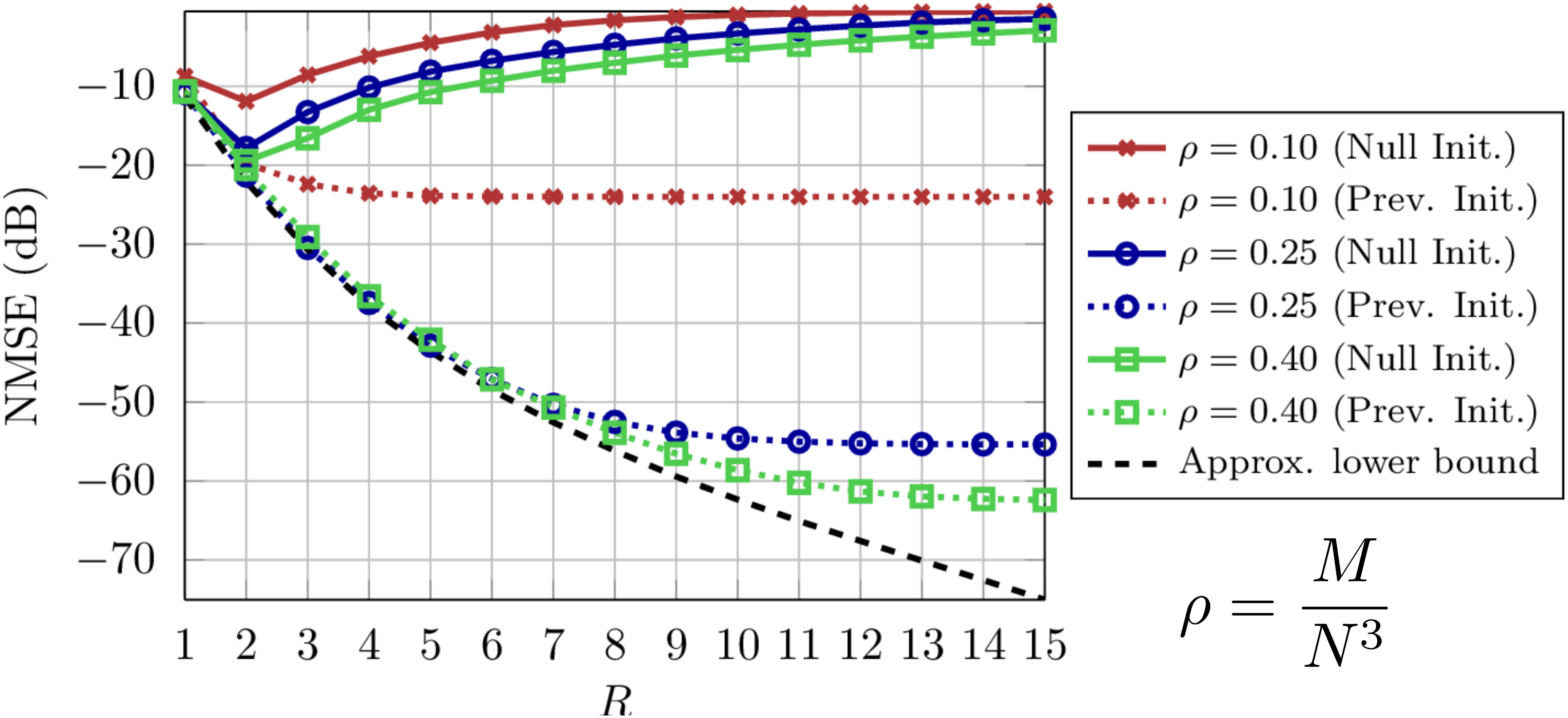
- Recovery of 20x20x20 T2 tensors



$$\rho = \frac{M}{N^3}$$

Recovery performance (TC)

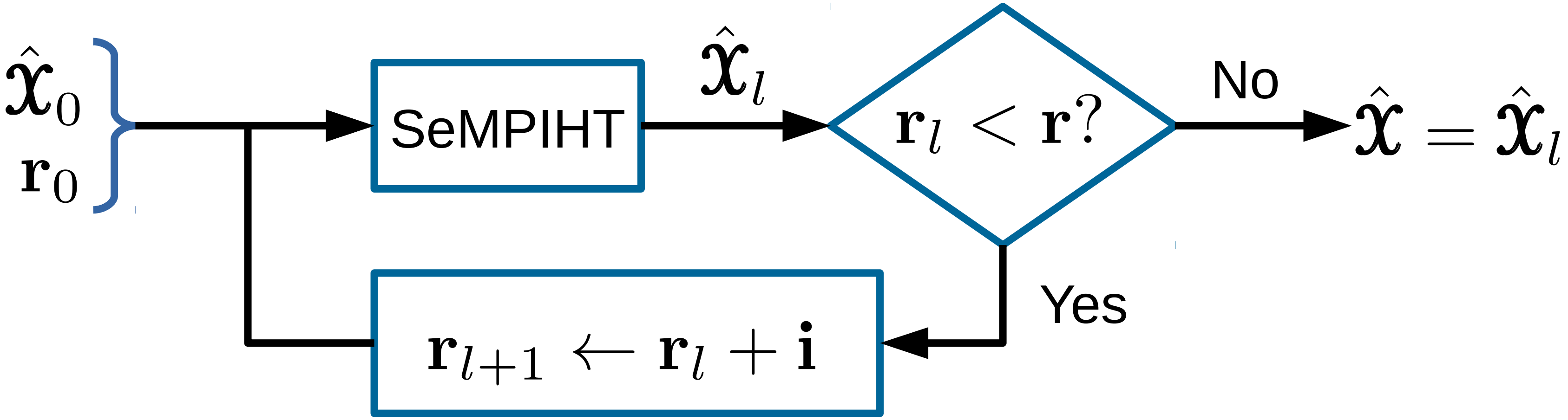
- Recovery of T2 tensors
  - Non-ideal coherence properties



$$\rho = \frac{M}{N^3}$$

SeMPIHT with gradual rank increase (GRI)

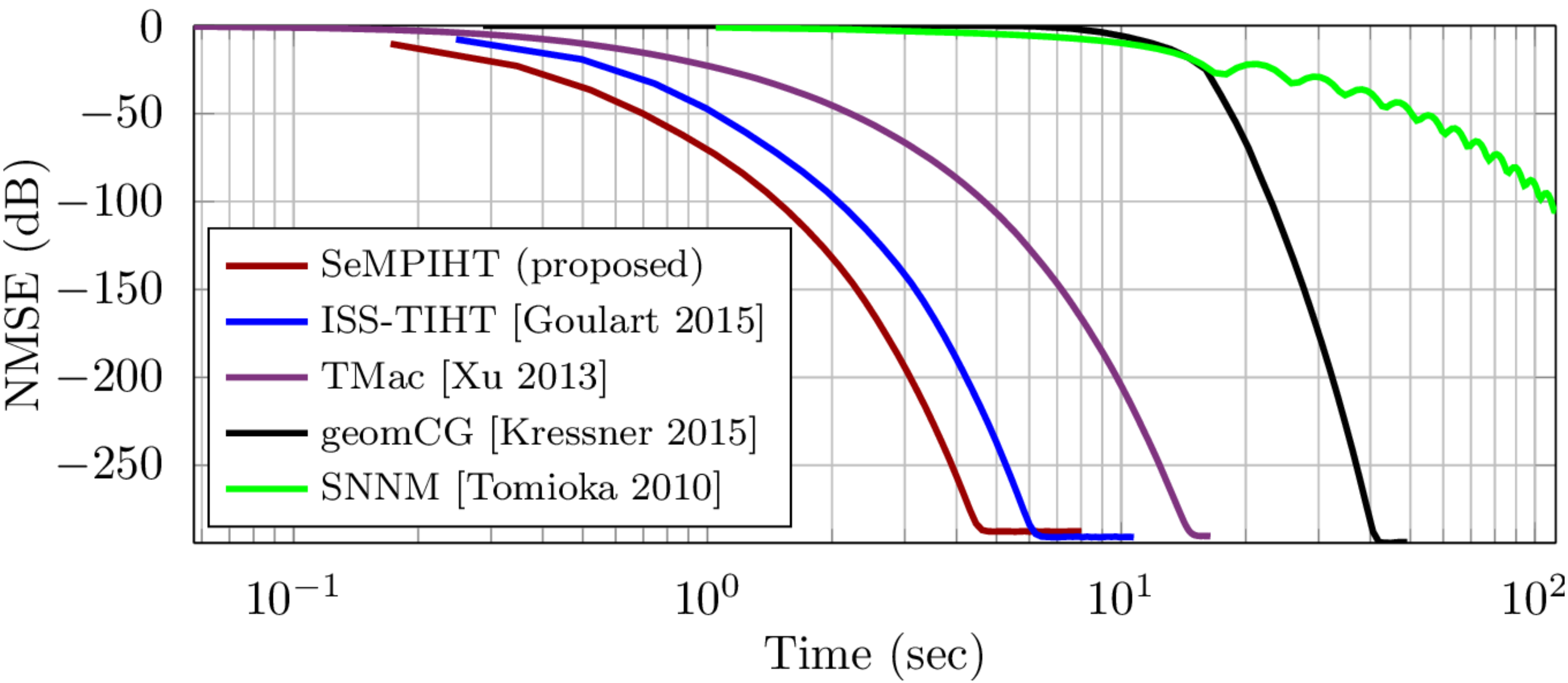
- Starts off with low mrank components $\mathbf{r}_0$
- Runs SeMPIHT, increments components and repeats, until attaining the target mrank $\mathbf{r}$



- Continuation scheme yielding increasingly complex intermediate solutions
- Accelerates convergence
- Avoids degradation due to non-ideal coherence properties
- Only makes sense for decaying spectra (T2 tensors)

Convergence speed (TC)

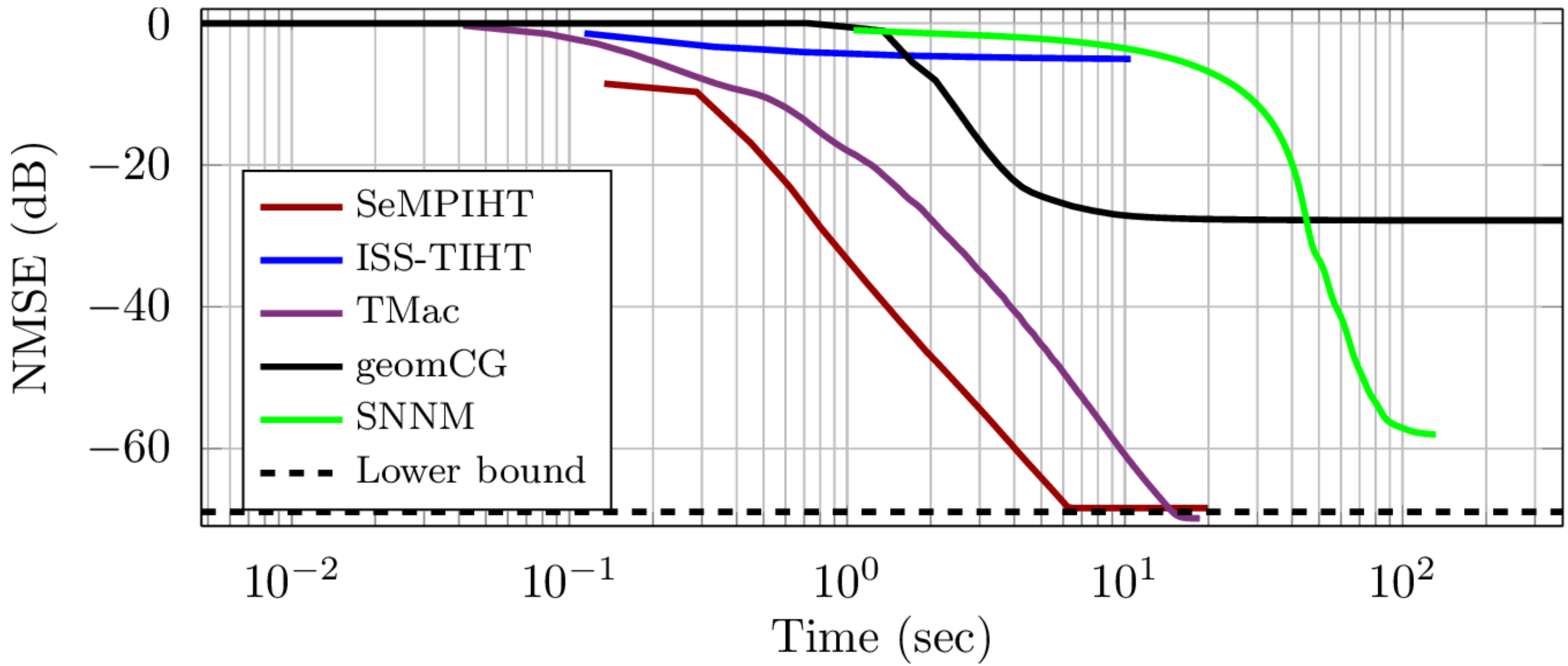
- 100x100x100 T1 tensors (without GRI)
- $\rho = 0.2, \quad \mathbf{r} = (10, 10, 10)$



- Cost reduction due to SeMP

Convergence speed (TC)

- 100x100x100 T2 tensors
- $\rho = 0.2, \quad \mathbf{r} = (30, 30, 30)$



- GRI allows escaping local minima

Concluding remarks

- SeMP is less costly than truncated HOSVD, having superior or comparable performance in IHT
- Sequential optimality of projections enables deriving performance bounds
- Yet, implied sampling bounds are suboptimal
  - Observed : optimal for Gaussian MOs
- GRI improves convergence speed, stabilizes error when model is overcomplex and copes with non-ideal coherence in TC

References

- [Cheng 2016] H. Cheng *et al*. "Scalable and Sound Low-Rank Tensor Learning." Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. 2016.
- [De Lathauwer 2000] L. De Lathauwer, B. De Moor and J. Vandewalle. "A multilinear singular value decomposition." SIAM journal on Matrix Analysis and Applications 21.4 (2000): 1253-1278.
- [Goulart 2015] J. H. M. Goulart and G. Favier. "An iterative hard thresholding algorithm with improved convergence for low-rank tensor recovery." EUSIPCO, 2015.
- [Goulart 2016] J. H. M. Goulart and G. Favier. "Low-rank tensor recovery using sequentially optimal modal projections in iterative hard thresholding (SeMPIHT)", 2016 (submitted).
- [Huang 2014] B. Huang *et al*. "Provable low-rank tensor recovery." Optimization-Online 4252 (2014): 2.

References

- [Kressner 2014] D. Kressner, M. Steinlechner and B. Vandereycken. "Low-rank tensor completion by Riemannian optimization." BIT Numerical Mathematics 54.2 (2014): 447-468.
- [Mu 2014] C. Mu *et al*. "Square Deal: Lower Bounds and Improved Relaxations for Tensor Recovery." ICML. 2014.
- [Rauhut 2013] H. Rauhut, R. Schneider and Z. Stojanac. "Low rank tensor recovery via iterative hard thresholding". 10th Int. Conf. Sampling Theory Applicat., 2013.
- [Rauhut 2016] H. Rauhut, R. Schneider and Z. Stojanac. "Low rank tensor recovery via iterative hard thresholding." arXiv preprint arXiv:1602.05217 (2016).

References

- [Tomioka 2010] R. Tomioka, K. Hayashi and H. Kashima. "Estimation of low-rank tensors via convex optimization." arXiv preprint arXiv:1010.0789 (2010).
- [Tomioka 2011] R. Tomioka *et al*. "Statistical performance of convex tensor decomposition." Advances in Neural Information Processing Systems. 2011.
- [Vannieuwenhoven 2012] N. Vannieuwenhoven, R. Vandebril and K. Meerbergen. "A new truncation strategy for the higher-order singular value decomposition." SIAM Journal on Scientific Computing 34.2 (2012): A1027-A1052.
- [Xu 2013] Y. Xu *et al*. "Parallel matrix factorization for low-rank tensor completion." arXiv preprint arXiv:1312.1254 (2013).
- [Zhang 2015] M. Zhang, L. Yang and H. Zheng-Hai. "Minimum n-rank approximation via iterative hard thresholding." Applied Mathematics and Computation 256 (2015): 860-875.