Tensor Sparsity and Near-Minimal Rank Approximation for High-Dimensional PDEs

Wolfgang Dahmen, RWTH Aachen

Collaborators: Markus Bachmayr, Ron DeVore, Lars Grasedyck, Endre Süli

Paris, Oct.11, 2013

Contents

High-Dimensional Problems

- Data mining
- Parameter dependent PDEs: $d =$ spatial dim. + parameter dim.
- Stochastic PDEs: $d = \infty$
- Electronic Schrödinger equation: $d = 3N$
- Fokker-Planck equations for polymeric fluids:
  $d = 3K,\ K =$ length of polymer chains

Curse of Dimensionality: "intractability" results (Novak/Woźniakowski)

$$\text{Accuracy } \epsilon \quad \leftrightarrow \quad \text{comp. cost } \sim \epsilon^{-d/s} \quad \rightsquigarrow$$

Remedies (?):

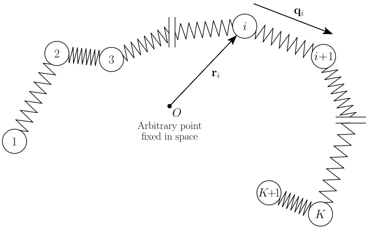
- "Excessive" regularity
- "Hidden sparsity" with respect to a problem dependent dictionary
  ... separation of variables...

Main Paradigms

Parameter dependent PDEs: (e.g. Reduced Basis Method) [Maday, Patera,...BCDDGW]

$$a(u,v;p) = \langle f, v\rangle, \quad v \in X,\ p \in \mathcal{P} \subset \mathbb{R}^m, \quad \rightsquigarrow\ u(x,p) = \sum_{i=1}^{n} c_i(p) u(x, p_i)$$

High dimensional phase space: e.g. Fokker-Planck eqs. ...



...operator splittings ... $\rightsquigarrow$ ... high-dimensional diffusion equation on product domain [Barrett/Süli]

Products help...

$f \in C^s([0,1]^d)$

$$f(x) = f(x_1, \ldots, x_d) \quad \approx \quad \sum_{k=1}^{r} f_{k,1}(x_1) \cdots f_{k,d}(x_d)$$

$$\approx \sum_{|\nu|_\infty \le n} c_\nu \psi_\nu(x) \qquad \sum_{k=1}^{r} \prod_{\ell=1}^{d} \Big(\underbrace{\sum_{j \le n} c_{\ell,j} \psi_j(x_\ell)}_{\approx f_{k,\ell}(x_\ell)} \Big)$$

$d.o.f.:$ $\quad \sim n^d =: N \qquad rdn =: N$

accur.: $\quad O(n^{-s}) = O(N^{-s/d}) \qquad rdn^{-s} = rd^{1+s} N^{-s}$

$work/acc.:$ $\quad N \sim \varepsilon^{-d/s} \qquad N \sim r^{1/s} d^{1+1/s} \varepsilon^{-1/s}$

Setting: e.g. $\mathcal{D} = -\sum_{j=1}^{d} \partial_{x_j}^2 = -\Delta$

$$\mathcal{D} = \sum_{i=1}^{d} \mathcal{I}_1 \otimes \cdots \otimes \mathcal{I}_{i-1} \otimes \mathcal{D}_i \otimes \mathcal{I}_{i+1} \otimes \cdots \otimes \mathcal{I}_d,$$

$\mathcal{D}_j : H_j(\Omega_j) \to (H_j(\Omega_j))'$ H_j-elliptic, $\times_{j=1}^{d} \Omega_j =: \Omega \subset \mathbb{R}^{dp}$

$$\mathrm{H} := \bigcap_{j=1}^{d} \{ L_2(\Omega_1) \otimes \cdots \otimes L_2(\Omega_{j-1}) \otimes H_j(\Omega_j) \otimes L_2(\Omega_{j+1}) \otimes \cdots \otimes L_2(\Omega_d) \}$$

$$\|v\|_{\mathrm{H}}^2 := \langle \mathcal{D}v, v \rangle , \quad a(u,v) := \langle \mathcal{D}u, v \rangle \ \forall\, v, w \in \mathrm{H}$$

$$\rightsquigarrow \mathcal{D} : \mathrm{H} \to \mathrm{H}', \quad \mathrm{H} \subset L_2(\Omega) \subset \mathrm{H}', \quad \|v\|_{\mathrm{H}^s}^2 = \|v\|_s^2 := \langle \mathcal{D}^s v, v \rangle , \ \mathrm{H} = \mathrm{H}^1$$

Solution structure (?)

$$a(u,v) = \langle f, v \rangle , \quad v \in \mathrm{H}$$

Main Objectives

"Regularity":

$a(u, v) = \langle f, v\rangle$, $v \in \mathrm{H}$ ... suppose that f is tensor sparse...

$$\ldots \overset{?}{\Longrightarrow} \; u = \mathcal{D}^{-1} f \text{ is also tensor sparse?} \quad u \approx \sum_{k=1}^{n(\varepsilon)} u_{k,1} \otimes \cdots \otimes u_{k,d}$$

Computatibility:

Compute tensor sparse appr's to $\mathcal{D}^{-1} f$ with "near-minimal cost"

- realize (near-)minimal ranks
- find (near-)optimally sparse representations of tensor factors

Issues:

- stability of tensor formats [Lathauwer, Hackbusch, Grasedyck, Oseledets,...]
- continuous versus discrete ... a scaling trap

Tensor Sparsity

Model 1:

$$\Sigma_n := \left\{ g = \sum_{k=1}^{r} \bigotimes_{j=1}^{d} g_{k,j} \,:\, g_{k,j} = \sum_{\mu\in\Gamma_{k,j}} c_{k,j,\mu} e_{j,\mu},\ \sum_{k=1}^{r}\sum_{j=1}^{d} \#(\Gamma_{k,j}) \le n \right\}$$

$$\sigma_n(f)_{\mathrm{H}^t} := \inf\{\|f-g\|_{\mathrm{H}^t} : g \in \Sigma_n\}$$

Given a "growth sequence" $\gamma(n) \nearrow \infty,\ n \to \infty$

$$\mathcal{A}^{\gamma}_{(1)}((\Sigma_n), \mathrm{H}^t) := \{f \in \mathrm{H}^t : |f|_{\gamma,t} := \sup_{n\in\mathbb{N}} \gamma(n)\sigma_n(f)_{\mathrm{H}^t} < \infty\},$$

How to read this ...

$$v \in \mathcal{A}^{\gamma}_{(1)}((\Sigma_n), \mathrm{H}^t) \ \Rightarrow\ \sigma_n(v)_{\mathrm{H}^t} \le \gamma(n)^{-1}|v|_{\gamma,t} \sim \varepsilon$$

$$\rightsquigarrow\ \exists\, v_\varepsilon \in \Sigma_{\gamma^{-1}(|v|_{\gamma,t}/\varepsilon)}, \quad \text{such that} \quad \|v - v_\varepsilon\|_t \le \varepsilon$$

...it takes $\gamma^{-1}(|v|_{\gamma,t}/\varepsilon)$ d.o.f./rank to achieve accuracy ε in H^t

Tensor Sparsity

Model 2:

$$\Sigma_n^{s,b} := \left\{ g = \sum_{k=1}^{n} \bigotimes_{j=1}^{d} g_{k,j} : g_k \in \mathrm{H}^s,\ \|g_k\|_s, \|g\|_s \le b \right\}, \quad \Sigma_n := \bigcup_{b>0} \Sigma_n^{s,b}$$

$$t < s : \qquad \sigma_n^{s,b}(f)_{\mathrm{H}^t} := \inf_{g \in \Sigma_n^{s,b}} \|f - g\|_t$$

$$|f|_{\gamma,b,t,s} := \sup_{n \in \mathbb{N}} \gamma(n) \sigma_n^{s,b}(f)_{\mathrm{H}^t}$$

$$\mathcal{A}_{(2)}^{\gamma}((\Sigma_n), \mathrm{H}^t) := \{ f \in \mathrm{H}^t : \exists\, b < \infty, \text{ s.t. } |f|_{\gamma,b,t,s} < \infty \}$$

Sparsity Results

Theorem 1:

$$i \in \{1,2\}: \quad f \in \mathcal{A}^{\gamma}_{(i)}((\Sigma_n), \mathrm{H}^{-1+\zeta}) \Rightarrow u = \mathcal{D}^{-1} f \in \mathcal{A}^{\hat{\gamma}}{}_{(i)}((\Sigma_n), \mathrm{H}^1)$$

where

$$\hat{\gamma}(n) := (\gamma \circ G^{-1})(n), \quad G(n) := \kappa(\zeta) n \big(\log(\gamma(n) C(f))\big)^2$$

Complexity:

- $i = 1$: $|u|_{\hat{\gamma},1} \leq 2|f|_{\gamma,-1+\zeta}$, accur. $\varepsilon \Rightarrow \#(\mathbf{c}(u_\varepsilon)) \leq \hat{\gamma}^{-1}\big(2|f|_{\gamma,-1+\zeta}/\varepsilon\big)$
- $i = 2$: $\#(\mathbf{c}(u_\varepsilon)) \leq \big(C_1(f,\zeta)\gamma^{-1}(\varepsilon^{-1})(\log(\varepsilon))^2 d\big)^{1+1/\zeta} \varepsilon^{-1/\zeta}$

  $\#\mathbf{ops}(u_\varepsilon) = O\Big(d^{1+\frac{2}{\zeta}} \log(dF(\varepsilon)) F(\varepsilon)^{\frac{2}{\zeta}}\Big)$, $F(\varepsilon) := \frac{\kappa\gamma^{-1}\left(\frac{C}{\varepsilon}\right)\left(\log(\varepsilon)\right)^2}{\varepsilon}$

Examples:

- $i = 1$: $\gamma(n) = n^\alpha \rightsquigarrow \hat{\gamma}(n) \sim (n/C \log n)^\alpha$
- $i = 2$: $\gamma(n) = e^{\alpha n} \rightsquigarrow \hat{\gamma}(n) \sim e^{C(\zeta,f)(\alpha n)^{1/3}}$

A Tool: Exponential Sums ...[Braess/Hackbusch]

$$\sup_{x\in[1,\infty]} |x^{-1} - s_r^*(x)| \le Ce^{-\pi\sqrt{r}}, \quad s_r^*(x) = \sum_{k=1}^{r} \omega_{r,k} e^{-\alpha_{r,k} x}$$

$\tau = \tau_1 \otimes \cdots \otimes \tau_d$

$$\mathcal{D}^{-1}\tau \quad \approx \quad s_r^*(\mathcal{D})\tau := \sum_{k=1}^{r} \omega_{r,k} e^{-\alpha_{r,k}\mathcal{D}}\tau = \sum_{k=1}^{r} \omega_{r,k} \Big(\bigotimes_{j=1}^{d} \underbrace{e^{-\alpha_{r,k}\mathcal{D}_j}\tau_j}_{g_j} \Big)$$

PROPOSITION 1:

For $-1 \le t \le s \le 1$, $\tau \in \mathrm{H}^t$, one has

$$\|\mathcal{D}^{-1} - s_r^*(\mathcal{D})\|_{\mathrm{H}^t \to \mathrm{H}^s} \le Ce^{-\frac{(2-s+t)\pi}{2}\sqrt{r}}$$

- Eigensystem for the $\mathcal{D}_j$: $\{e_{j,k}\}_{k\in\mathbb{N}}, \mathcal{D}_j e_{j,k} = \lambda_{j,k} e_{j,k}$
- $\rightsquigarrow e_\nu := e_{1,\nu_1} \otimes \cdots \otimes e_{d,\nu_d}$, $\mathcal{D}e_\nu = \lambda_\nu e_\nu$, $\lambda_\nu = \lambda_{1,\nu_1} + \cdots + \lambda_{d,\nu_d}$
- $\|v\|_s^2 = \sum_{\nu\in\mathbb{N}^d} \lambda_\nu^s |\langle v, e_\nu\rangle|^2$

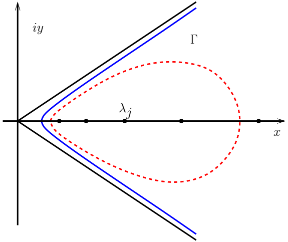
Complexity

cost$(\mathcal{D}^{-1}f, \varepsilon)$:= computational cost of solving $\mathcal{D}u = f$ with accuracy ε

- $f = \tau$: evaluate exponentials

$$s_r^*(\mathcal{D})\tau = \sum_{k=1}^{r} \omega_{r,k}\Big(\bigotimes_{j=1}^{d} e^{-\alpha_{r,k}\mathcal{D}_j}\tau_j\Big), \quad r = r(\varepsilon) \sim |\log \varepsilon|^2$$

$$e^{-t\mathcal{D}_j}\tau_j = \frac{1}{2\pi i}\int_\Gamma e^{-t\gamma}(\gamma I - \mathcal{D}_j)^{-1}\tau_j d\gamma,$$



...truncation, sinc-quadrature $\rightsquigarrow$
$d|\log \varepsilon|$ solves at cost $(\varepsilon/d)^{-1/\zeta}$ for $|\log \varepsilon|^2$ terms $\rightsquigarrow$

- **cost**$(\mathcal{D}^{-1}\tau, \varepsilon) \lesssim d^{1+1/\zeta}\varepsilon^{-1/\zeta}|\log \varepsilon|^3$ (instead of: $\varepsilon^{-d/\zeta}$)

Inventory

Some facts: [de Silva, Lathauwer, Hackbusch, Grasedyck, Oseledets, Schneider...]

- "canonical format" $\sum_{k\in\mathbb{N}} u_{k,1} \otimes \cdots \otimes u_{k,d}$ i.g. unstable
- optimal subspace methods: unique best approximation exists and is realized by orthogonal projections - T- /(H-T)-formats
- HOSVD $\rightsquigarrow$ near minimimal rank approximation
- efficient numerical tools [Espig, Kolda,...]

Operator equations:

- immediate reduction to a fixed discrete system
- accuracy considerations detached from continuous solution
- approximation error and residuals are measured in the same (Euclidean) norm - "scaling trap"
- accuracy and rank growth cannot be controlled simultaneously
- PGD...convergence, ranks?... [Falcó, Chinesta, Nouy,...]

Reduction to Problem in ℓ_2

"Universal background" basis:

$$\{\psi_\nu = \psi_{\nu_1} \otimes \cdots \otimes \psi_{\nu_d} : \nu \in \mathcal{J}^d\} \text{ O.N.B. for } L_2(\Omega) \rightsquigarrow$$

$$\Psi = \Big\{ \Big(\sum_{i=1}^{d} 2^{2|\nu_i|} \Big)^{-\frac{1}{2}} \psi_\nu =: s_\nu \psi_\nu \Big\}_{\nu \in \mathcal{J}^d} \text{ Riesz-basis for } \mathrm{H} \subset L_2(\Omega)$$

$$\mathcal{D}u = f \Leftrightarrow \mathbf{A}\mathbf{u} = \mathbf{f}, \quad \mathbf{A} = \big(s_\nu a(\psi_\nu, \psi_\mu) s_\mu\big)_{\nu,\mu \in \mathcal{J}}, \ \mathbf{f} = \big(\langle f, s_\nu \psi_\nu \rangle\big)_{\nu \in \mathcal{J}}$$

Theorem:

$$\kappa(\mathbf{A}) := \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \lesssim 1$$

$$u \in \mathrm{H} \quad \leftrightarrow \quad \mathbf{u} = (u_\nu)_{\nu \in \mathcal{J}^d} \in \ell_2(\mathcal{J}^d)$$

Scheme: Perturbed Ideal Iteration

Algorithm:

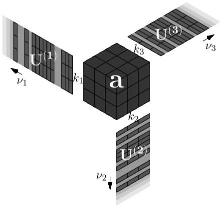
$$\mathbf{u}^{k+1} = \mathrm{C}_{\varepsilon_3(k)}\big(\mathrm{P}_{\varepsilon_2(k)}(\mathbf{u}^k + \omega(\mathbf{f} - \mathbf{A}\mathbf{u}^k))\big) \rightsquigarrow \|\mathbf{u} - \mathbf{u}^{k+1}\| \le \rho\|\mathbf{u} - \mathbf{u}^k\|, \ \rho < 1$$

Mode frames: $\mathbf{U}_k^{(j)} \in \ell_2(\mathcal{J}),\ k \in \mathbb{N},\ j = 1, \dots, d,\ \langle \mathbf{U}_k^{(i)}, \mathbf{U}_l^{(i)} \rangle = \delta_{kl},\ k, l \in \mathbb{N}$

Tucker format:

$$\mathbf{u} = \sum_{k_1=1}^{\infty} \cdots \sum_{k_d=1}^{\infty} \langle \mathbf{u}, \mathbf{U}_{k_1}^{(1)} \otimes \cdots \otimes \mathbf{U}_{k_d}^{(d)} \rangle\, \mathbf{U}_{k_1}^{(1)} \otimes \cdots \otimes \mathbf{U}_{k_d}^{(d)} =: \sum_{\mathbf{k} \in \mathbb{N}^d} \mathbf{c}_\mathbf{k} \mathbb{U}_\mathbf{k}$$

Hierarchical Tucker (H-T)-format:
hierarchical factorization of
core tensor $(\mathbf{c}_\mathbf{k})_{\mathbf{k} \in \mathbb{N}^d}$



Some New Ingredients...

- Wavelet techniques for the $1D$-tensor factors (coarsening, best N-term approximation)
- Thresholding Lemma: restoring a near-minimal rank approximation to the unknown solution from given approximations
- Contractions:

$$\pi^{(i)}(\mathbf{u}) = \left(\pi^{(i)}_{\nu_i}(\mathbf{u})\right)_{\nu_i\in\mathcal{J}} := \left(\Big(\sum_{\check{\nu}_i} |u_\nu|^2\Big)^{\frac{1}{2}}\right)_{\nu_i\in\mathcal{J}} \quad \rightsquigarrow$$

$$\pi^{(i)}_\nu(\mathbf{u}) = \Big(\sum_k \big|\mathbf{U}^{(i)}_{\nu,k}\big|^2 \big|\sigma^{(i)}_k\big|^2\Big)^{\frac{1}{2}}, \quad \pi^{(i)}_\nu(P_{\mathbb{U}(\mathbf{u}),\mathbf{r}}\mathbf{u}) \le \pi^{(i)}_\nu(\mathbf{u}), \quad \nu\in\mathcal{J}$$

- Exponetial sum approximation to (non-separable) scaling matrices $\mathbf{S} = (s_\nu\delta_{\nu,\mu})_{\nu,\mu\in\mathcal{J}^d}$ in $\boldsymbol{A} = \mathbf{STS}$

Optimal Convergence

Benchmarks/Assumptions (cf. Model 2):

- $\mathbf{u}$ is tensor sparse $\mathbf{u} \in \mathcal{A}^{\gamma_{\mathbf{u}}}((\Sigma_n^{HT}), \ell_2(\mathcal{J}^d)) =: \mathcal{A}_{HT}^{\gamma_{\mathbf{u}}}$
- $\boldsymbol{A}$ is "tensor sparse" - can be well approximated by low rank matrices
- $\pi^{(i)}(\mathbf{u}) \in \mathcal{A}^s$, $i \leq d$, $(\mathbf{v} \in \mathcal{A}^s \Leftrightarrow \sup_n n^s(\inf_{\mathrm{supp}\, \mathbf{z} \leq n} \|\mathbf{v} - \mathbf{z}\|) =: |\mathbf{v}|_{\mathcal{A}^s} < \infty)$
- The low-rank approximations to $\boldsymbol{A}$ are s^*-compressible with $s^* > s$

Theorem 2: For $\varepsilon > 0$ the **Algorithm** produces a $\mathbf{u}_\varepsilon$ with $\|\mathbf{u} - \mathbf{u}_\varepsilon\| \leq \varepsilon$ s.t.:

$$|\mathrm{rank}\, \mathbf{u}_\varepsilon|_\infty \leq \gamma_{\mathbf{u}}^{-1}(C\|\mathbf{u}\|_{\mathcal{A}_{HT}^{\gamma_{\mathbf{u}}}}/\varepsilon), \quad \sum_{i=1}^d \#(\mathrm{supp}_i(\mathbf{u}_\varepsilon)) \lesssim \Big(\sum_{i=1}^d \|\pi^{(i)}(\mathbf{u})\|_{\mathcal{A}^s}/\varepsilon\Big)^{1/s}$$

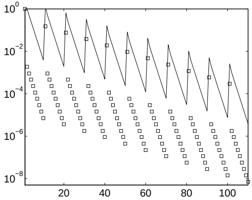
$\mathrm{supp}_i(\mathbf{u}) := \bigcup_{k\in\mathbb{N}} \mathrm{supp}\, \mathbf{U}_k^{(i)}$. Stability in $\mathcal{A}_{HT}^{\gamma_{\mathbf{u}}}, \mathcal{A}^s$:

$$\|\mathbf{u}_\varepsilon\|_{\mathcal{A}_{HT}^{\gamma_{\mathbf{u}}}} \leq C\|\mathbf{u}\|_{\mathcal{A}_{HT}^{\gamma_{\mathbf{u}}}}, \quad \sum_{i=1}^d \|\pi^{(i)}(\mathbf{u}_\varepsilon)\|_{\mathcal{A}^s} \lesssim \sum_{i=1}^d \|\pi^{(i)}(\mathbf{u})\|_{\mathcal{A}^s}$$

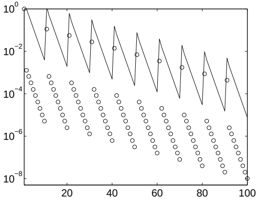
$$\#(\mathbf{ops}) \lesssim |\log \varepsilon|^{C(\boldsymbol{A},\mathbf{f},\log d)} \Big(\sum_{i=1}^d \max\{\|\pi^{(i)}(\mathbf{u})\|_{\mathcal{A}^s}, \|\pi^{(i)}(\mathbf{f})\|_{\mathcal{A}^s}\}/\varepsilon\Big)^{\frac{1}{s}}$$

Numerical Experiments

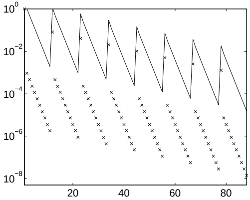
$$(Tv)(t) := \int_0^t v\,ds, \quad \Big(\mathfrak{I} - \omega_d \bigotimes_{i=1}^{d} T\Big)u = f, \quad f = \bigotimes_{i=1}^{d} \sqrt{2\pi}\,\chi_{[0,1/\pi]} \cos(2\pi^2\,\cdot)$$



$d = 32$

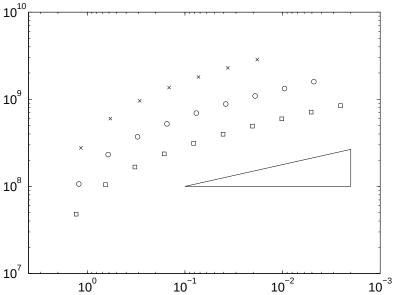


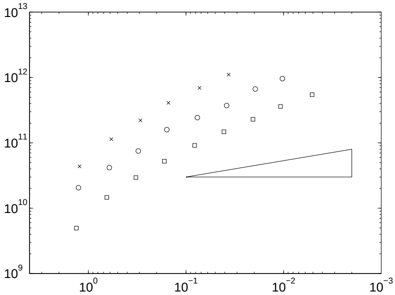
$d = 64$

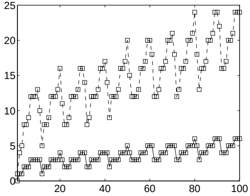


$d = 128$

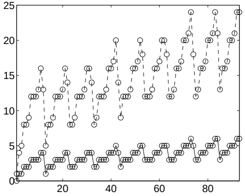
Numerical Experiments



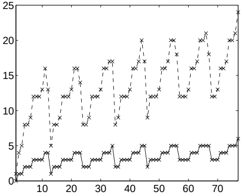




$d = 32$



$d = 64$



$d = 128$