Multi-Linear Mappings, SVD, HOSVD, and the Numerical Solution of Ill-Conditioned Tensor Least Squares Problems

Lars Eldén
Department of Mathematics, Linköping University

1 April 2005
Multi-Linear Integral Equations of the First Kind

\[ \int \int_{R_1} k(x_1, x_2, x_3, x_4) f(x_3, x_4) dx_3 dx_4 = g(x_1, x_2), \quad (x_1, x_2) \in R_2, \]

$R_1$ and $R_2$ are rectangular domains in $\mathbb{R}^2$.

The kernel function is smooth, e.g. $k \in L_2[R_1 \times R_2]$.

Ill-posed problem: the solution does not depend continuously on the data.

Applications: restoration of blurred images, inverse problems, ...
Discretization

\[ \sum_{1 \leq k, l \leq n} k_{ijkl} f_{kl} = g_{ij}, \quad 1 \leq i, j \leq n, \]

Corresponding least squares problem:

\[
\min_{f_{kl}} \sum_{1 \leq i, j \leq n} \left( \sum_{1 \leq k, l \leq n} k_{ijkl} f_{kl} - g_{ij} \right)^2.
\]

For simplicity: all subscripts \( i, j, k, l \) run from 1 to \( n \).

Ill-posedness \( \implies \) the linear system is extremely ill-conditioned (if \( n \) is large enough).
Tensor $\rightarrow$ Matrix

Stack the columns:

$$F \rightarrow \text{vec}(F) = f = \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot n} \end{pmatrix}, \quad G \rightarrow \text{vec}(F') = g = \begin{pmatrix} g_{\cdot 1} \\ \vdots \\ g_{\cdot n} \end{pmatrix}$$

Organize the coefficients $k_{ijkl}$ accordingly.

Standard linear system

$$Kf = g, \quad K \in \mathbb{R}^{n^2 \times n^2}.$$
Numerical Solution: SVD

\[ K f = g, \quad K \in \mathbb{R}^{n^2 \times n^2}. \]

Compute the singular value decomposition (SVD) of \( K \):

\[ K = U \Sigma V^T = \sum_{i=1}^{n^2} \sigma_i u_i v_i^T, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0. \]

Formal solution:

\[ f = \sum_{i=1}^{n^2} \frac{u_i^T g}{\sigma_i} v_i \]

\( \sigma_n \approx 0 \implies \) unstable: small perturbations of data cause large errors.
Truncated SVD: TSVD

Stabilize by truncating the expansion, $k << n^2$:

$$f_k = \sum_{i=1}^{k} \frac{u_i^T g}{\sigma_i} v_i$$

Computation of the SVD of $K$: $Cn^6$ flops, $C \approx 5$

Prohibitively costly if $n$ is large.

Alternative: Iterative methods (sparsity)

Tensor problem: Is it possible to do better using TENSOR methods?
Basic Tensor Concepts

Mode—$I$ multiplication of a tensor by a matrix

\[(A \times_1 U)(j, i_2, i_3, i_4) = \sum_{i_1=1}^{n} a_{i_1,i_2,i_3,i_4} u_{j,i_1}.\]

All column vectors in the 4-tensor are multiplied by the matrix $U$.

Mode—$2$ multiplication by a matrix $V$: all row vectors are multiplied by the matrix $V$.

Mode—$3$ and mode—$4$ multiplication are analogous.
\[ K \neq M, \text{ mode}-K \text{ and mode}-M \text{ multiplication commute:} \]

\[ A \times_K U \times_M V = A \times_M V \times_K U. \]

Useful identity:

\[ A \times_K F \times_K G = A \times_K GF. \]
Matricizing the Tensor

Matricizing of a 3–mode tensor $\mathcal{A}$:

\[
\mathbb{R}^{n \times n^2} \ni A_{(1)} = \begin{pmatrix}
\mathcal{A}(;1;) & \mathcal{A}(;2;) & \cdots & \mathcal{A}(;n;)
\end{pmatrix},
\]

\[
\mathbb{R}^{n \times n^2} \ni A_{(2)} = \begin{pmatrix}
\mathcal{A}(;::, 1)^T & \mathcal{A}(;::, 2)^T & \cdots & \mathcal{A}(;::, n)^T
\end{pmatrix},
\]

\[
\mathbb{R}^{n \times n^2} \ni A_{(3)} = \begin{pmatrix}
\mathcal{A}(1, ::)^T & \mathcal{A}(2, ::)^T & \cdots & \mathcal{A}(n, ::)^T
\end{pmatrix}.
\]

Mode-$M$ multiplication $B = \mathcal{A} \times_M M$:

\[
B_{(M)} = MA_{(M)},
\]

followed by a reorganization of $B_{(M)}$ to the tensor $\mathcal{B}$.
Inner Product

Two tensors $\mathcal{A}$ and $\mathcal{B}$ of the same dimensions:

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1, i_2, i_3, i_4} a_{i_1 i_2 i_3 i_4} b_{i_1 i_2 i_3 i_4}.$$ 

Special case of contracted product of two tensors

The linear system $\sum_{1 \leq k, l \leq n} k_{ijkl} f_{kl} = g_{ij}, \quad 1 \leq i, j \leq n,$

$$\langle \mathcal{K}, \mathcal{F} \rangle_{\{3,4;1,2\}} = G,$$

The matrices $F$ and $G$ are identified with tensors $\mathcal{F}$ and $\mathcal{G}.$
Least squares problem

$$\|\langle \mathcal{K}, F \rangle_{\{3,4;1,2\}} - G \|,$$

$\| \cdot \|$ is the matrix Frobenius norm, or, equivalently, the tensor Frobenius norm,

$$\|G\| = \|\mathcal{G}\| = \langle \mathcal{G}, \mathcal{G} \rangle^{1/2}.$$
Some Obvious Facts

The tensor Frobenius norm is invariant under mode–$M$ multiplication by an orthogonal matrix:

**Proposition 1.** Let $U \in \mathbb{R}^{n \times n}$ be orthogonal. Then

$$\| \mathcal{A} \times_M U \| = \| \mathcal{A} \|. $$
Tensor-matrix equivalent of the identity $SV^Tx = Sy$, for $y = V^Tx$.

**Proposition 2.** Assume that $S$ is a 4–tensor, $U^{(3)}$ and $U^{(4)}$ are square matrices, and that $F$ is a matrix. Then

$$\langle S \times_3 U^{(3)} \times_4 U^{(4)}, F \rangle_{3,4;1,2} = \langle S, \tilde{F} \rangle_{3,4;1,2},$$

where

$$\tilde{F} = (U^{(3)})^TFU^{(4)}.$$
Matrix case: \( USx = U(Sx) \)

**Proposition 3.**

\[
\langle S \times_1 U^{(1)} \times_2 U^{(2)}, F \rangle_{\{3,4;1,2\}} = U^{(1)} \left( \langle S, F \rangle_{\{3,4;1,2\}} \right) (U^{(2)})^T.
\]
Tensor SVD (HOSVD)

The singular value decomposition of a 4–tensor\(^1\)

\[ \mathcal{A} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \times_4 U^{(4)}, \]

where \(U^{(i)} \in \mathbb{R}^{n \times n}\) are orthogonal matrices.

Core tensor \(\mathcal{S}\) has the same dimensions as \(\mathcal{A}\);

\(\mathcal{S}\) is all-orthogonal: slices along any mode are orthogonal. Let \(i \neq j\); then

\[
\langle \mathcal{S}(i,\cdot,\cdot,\cdot), \mathcal{S}(j,\cdot,\cdot,\cdot) \rangle = \langle \mathcal{S}(::,i,\cdot,\cdot), \mathcal{S}(::,j,\cdot,\cdot) \rangle \\
= \langle \mathcal{S}(::,\cdot,i,:), \mathcal{S}(::,\cdot,j,:) \rangle = \langle \mathcal{S}(::,\cdot,\cdot,\cdot,i), \mathcal{S}(::,\cdot,\cdot,\cdot,j) \rangle = 0.
\]

\(^1\)Equivalent to the Tucker-3 decomposition in psychometrics and chemometrics.
HOSVD

$$A = U^{(1)} S U^{(2)}$$
Singular Values

Mode–1 singular values

\[ \sigma^{(1)}_j = \|S(j,:,:)\|, \quad j = 1, \ldots, n. \]

The singular values are ordered,

\[ \sigma^{(i)}_1 \geq \sigma^{(i)}_2 \geq \cdots \geq \sigma^{(i)}_n \geq 0, \quad i = 1, 2, 3, 4. \]
The singular values are measures of the "energy" of the tensor

**Proposition 4.**

\[
\| A \|^2 = \| S \|^2 = \sum_{i=1}^{n} \left( \sigma_i^{(1)} \right)^2 = \cdots = \sum_{i=1}^{n} \left( \sigma_i^{(4)} \right)^2.
\]

The "energy" (mass) is concentrated at the \((1, 1, 1, 1)\) corner of the tensor

We can truncate the HOSVD (in analogy to TSVD)
Approximation

Proposition 5. Define a tensor \( \hat{A} \) by discarding the smallest singular values along each mode, \((\sigma_{i_1}^{(1)})_{i_1=k_1+1}^n, \ldots (\sigma_{i_4}^{(4)})_{i_4=k_4+1}^n\), i.e. setting the corresponding parts of the core tensor \( S \) equal to zero. Then the approximation error is bounded,

\[
\|A - \hat{A}\| = \|S - \hat{S}\| \leq \sum_{i_1=k_1+1}^{n} (\sigma_{i_1}^{(1)})^2 + \cdots + \sum_{i_4=k_4+1}^{n} (\sigma_{i_4}^{(4)})^2.
\]

No optimality property corresponding to Eckart-Young theorem for matrices.

The tensor \( \hat{A} \) defined in Proposition 5 is not the best rank-\( (k_1, \ldots, k_4) \) approximation of \( A \).
Computation of HOSVD

$U^{(i)}$ are the left singular matrices of $A^{(i)} \in \mathbb{R}^{n \times n^3}$.

1. for $j = 1, 2, 3, 4$
   
   (a) Compute the LQ decomposition $A^{(j)} = (L_j 0)Q_j$, without forming $Q_j$ explicitly.

   (b) Compute the SVD: $L_j = U^{(j)} \Sigma^{(j)} (V^{(j)})^T$, without forming $V^{(j)}$ explicitly.

2. Compute the core tensor:

   $$S = A \times_1 (U^{(1)})^T \times_2 (U^{(2)})^T \times_3 (U^{(3)})^T \times_4 (U^{(4)})^T$$
Computation

The mode—\( j \) singular values are the diagonal elements of \( \Sigma^{(j)} \):

\[
\Sigma^{(j)} = \text{diag}(\sigma_1^{(j)}, \sigma_2^{(j)}, \ldots, \sigma_n^{(j)}).
\]

Cost for computing HOSVD: \( Cn^5 \) flops.
Discrete Ill-Posed Problems

Linear Case: Truncated SVD Solution

\[ \min_f \|Kf - g\|_2, \quad K \in \mathbb{R}^{n^2 \times n^2}. \]

Solution (formally)

\[ f = V \Sigma^{-1} U^T g = \sum_{i=1}^{n^2} \frac{u_i^T g}{\sigma_i} v_i. \]
Solution for perturbed data, $\hat{g} = g + \delta$, (measurement or round off errors)

$$\hat{f} = \sum_{i=1}^{n^2} \frac{u_i^T \hat{g}}{\sigma_i} v_i = \sum_{i=1}^{n^2} \frac{u_i^T g}{\sigma_i} v_i + \sum_{i=1}^{n^2} \frac{u_i^T \delta}{\sigma_i} v_i.$$ 

The second term will explode.

Stabilization by TSVD:

$$f \approx \hat{f}_k = \sum_{i=1}^{k} \frac{u_i^T \hat{g}}{\sigma_i} v_i,$$
Multi-Linear Case: Truncated HOSVD

HOSVD

$$\mathcal{K} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \times_4 U^{(4)}.$$ 

Reduced least squares problem:

$$\min_{\tilde{F}} \| \langle \mathcal{S}, \tilde{F} \rangle_{\{3,4;1,2\}} - \mathcal{G} \|, \quad \tilde{F} = (U^{(3)})^T F U^{(4)}, \quad \mathcal{G} = (U^{(1)})^T G U^{(2)}.$$ 

Truncate the core tensor and solve:

$$\min_{\tilde{F}} \| \langle \tilde{\mathcal{S}}, \tilde{F} \rangle_{\{3,4;1,2\}} - \mathcal{G} \|,$$

where $\tilde{\mathcal{S}}(k + 1 : n, k + 1 : n, k + 1 : n, k + 1 : n) = 0$. 
Truncated HOSVD

\[ A \approx U_k^{(1)} S_k U_k^{(2)} U_k^{(3)} \]
Algorithm

1. Compute the HO singular values and vectors: $O(n^5)$ flops

2. Truncate to $k$ based on singular values, compute only $S_k$: $O(kn^4)$ flops

3. Solve

$$\min_{F_k} \| \langle S_k, F_k \rangle_{\{3,4;1,2\}} - G_k \|, \quad \iff \quad \min_{f_k} \| S_k f_k - g_k \|$$

(a) Compute SVD of $S_k$ and solve by TSVD: $O(k^6)$ flops
(b) Transform to original coordinates: $O(kn^2)$ flops
Numerical Example

Perturbation of data: $N(0, 1)$

Level of truncation of tensor: $k = 30$

Linear system level of truncation: $k_0 = 8$
Singular Values

![Singular Values Graph]
Unregularized Solution
Regularized Solution, $k = 8$
Error
Conclusions

- One order of magnitude faster than straightforward approach
- Relation HO singular values – Linear system singular values?
- Other applications: Tensor regression?