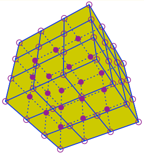


Cumulant Signal Processing, Tensors and some Recurring Problems

Phil Regalia

Department of Electrical Engineering and Computer Science
Catholic University of America
Washington, DC 20064

with thanks to E. Kofidis, M. Mboup, and V. S. Grigorascu

Future Directions in Tensors — NSF Feb 2009

Outline

1. Multilinear Product
2. An "Easy" Problem
3. Symmetric Case
4. Rank r Approximation
5. Displacement Structure

Multilinear Product

Given N matrices $\mathbf{A}^{(n)}$, the N-way (a.k.a. Tucker) product is

$$\mathcal{T}_{j_1,j_2,\ldots,j_N} = \mathbf{A}^{(1)} \star \mathbf{A}^{(2)} \star \cdots \star \mathbf{A}^{(N)} = \sum_{k=1}^{K} \mathbf{A}^{(1)}_{j_1,k} \mathbf{A}^{(2)}_{j_2,k} \cdots \mathbf{A}^{(N)}_{j_N,k}$$

With a core N-dimensional tensor $\mathcal{S}$, the weighted product is

$$\begin{aligned} \mathcal{T} &= \mathbf{A}^{(1)} \overset{\mathcal{S}}{\star} \mathbf{A}^{(2)} \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{A}^{(N)} \\ &= \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \cdots \sum_{k_N=1}^{K_N} \mathbf{A}^{(1)}_{j_1,k_1} \mathbf{A}^{(2)}_{j_2,k_2} \cdots \mathbf{A}^{(N)}_{j_N,k_N} \mathcal{S}_{k_1,k_2,\ldots,k_N} \end{aligned}$$

and is equivalent to

$$\mathrm{vec}(\mathcal{T}) = \left(\mathbf{A}^{(1)} \otimes \mathbf{A}^{(2)} \otimes \cdots \otimes \mathbf{A}^{(N)}\right) \mathrm{vec}(\mathcal{S})$$

Reduces to unweighted product when $\mathcal{S}_{k_1,k_2,\ldots,k_N} = \delta(k_1, k_2, \ldots, k_N)$.

An "Easy" Approximation Problem

Let $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$ be a collection of unit-norm column vectors.

Problem

Find unit-norm vectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$ and scalar σ to minimize Frobenius norm of

$$\mathcal{S} - \underbrace{\sigma \cdot \mathbf{u}^{(1)} \star \mathbf{u}^{(2)} \star \cdots \star \mathbf{u}^{(N)}}_{\text{rank one}}$$

Equivalent problem: Find unit norm vectors to maximize the functional

$$\begin{aligned} f\left(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}\right) &= (\mathbf{u}^{(1)})^T \overset{\mathcal{S}}{\star} (\mathbf{u}^{(2)})^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} (\mathbf{u}^{(N)})^T \\ &= \sum_{k_1} \sum_{k_2} \cdots \sum_{k_N} \mathbf{u}^{(1)}_{k_1} \mathbf{u}^{(2)}_{k_2} \cdots \mathbf{u}^{(N)}_{k_N} \mathcal{S}_{k_1, k_2, \ldots, k_N} \end{aligned}$$

An "Easy" Approximation Problem

Let $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$ be a collection of unit-norm column vectors.

Problem

Find unit-norm vectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$ and scalar σ to minimize Frobenius norm of

$$\mathcal{S} - \underbrace{\sigma \cdot \mathbf{u}^{(1)} \star \mathbf{u}^{(2)} \star \cdots \star \mathbf{u}^{(N)}}_{\text{rank one}}$$

Equivalent problem: Find unit norm vectors to maximize the functional

$$\begin{aligned} f\left(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}\right) &= (\mathbf{u}^{(1)})^T \overset{\mathcal{S}}{\star} (\mathbf{u}^{(2)})^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} (\mathbf{u}^{(N)})^T \\ &= \sum_{k_1} \sum_{k_2} \cdots \sum_{k_N} \mathbf{u}^{(1)}_{k_1} \mathbf{u}^{(2)}_{k_2} \cdots \mathbf{u}^{(N)}_{k_N} \mathcal{S}_{k_1, k_2, \ldots, k_N} \end{aligned}$$

Simple iterative scheme

Can write

$$f = \langle \nabla_1 f, \mathbf{u}^{(1)} \rangle,$$

Thus, given unit-norm vectors $\mathbf{u}^{(1,i)}, \ldots, \mathbf{u}^{(N,i)}$ at iteration i, update $\mathbf{u}^{(1)}$ as

$$\begin{aligned} \mathbf{v} &= \nabla_1 f = \sum_{k_2} \cdots \sum_{k_N} \mathbf{u}^{(2,i)}_{k_2} \cdots \mathbf{u}^{(N,i)}_{k_N} \mathcal{S}_{k_1,k_2,\ldots,k_N} \\ &= \mathbf{I} \overset{\mathcal{S}}{\star} (\mathbf{u}^{(2,i)})^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} (\mathbf{u}^{(N,i)})^T \\ \mathbf{u}^{(1,i+1)} &= \frac{\mathbf{v}}{\|\mathbf{v}\|} \end{aligned}$$

and likewise for $\mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \ldots, \mathbf{u}^{(N)}$, and repeat.

Each update gives an increase in $f(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)})$, so functional converges monotonically to a local maximum.

Symmetric case

Applications in blind deconvolution, source separation, or independent component analysis have a symmetric tensor:

$$\mathcal{S}_{k_1,k_2,\ldots,k_N} = \text{cum}\left(x_{k_1}, x_{k_2}, \ldots, x_{k_N}\right)$$

One seeks to maximize N-th order cumulant of $\mathbf{u}^T\mathbf{x}$ with $\|\mathbf{u}\| = 1$:

$$f(\mathbf{u}) = \text{cum}_N(\mathbf{u}^T\mathbf{x}) = \underbrace{\mathbf{u}^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{u}^T}_{N \text{ terms}}$$

Symmetric version of previous algorithm gives

$$\mathbf{v} = \sum_{k_2} \cdots \sum_{k_N} \mathbf{u}_{k_2}^{(i)} \cdots \mathbf{u}_{k_N}^{(i)} \mathcal{S}_{k_1,k_2,\ldots,k_N}$$

$$\mathbf{u}^{(i+1)} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Sometimes converges, sometimes not ...

Symmetric case

Applications in blind deconvolution, source separation, or independent component analysis have a symmetric tensor:

$$\mathcal{S}_{k_1,k_2,\ldots,k_N} = \text{cum}\left(x_{k_1}, x_{k_2}, \ldots, x_{k_N}\right)$$

One seeks to maximize N-th order cumulant of $\mathbf{u}^T\mathbf{x}$ with $\|\mathbf{u}\| = 1$:

$$f(\mathbf{u}) = \text{cum}_N(\mathbf{u}^T\mathbf{x}) = \underbrace{\mathbf{u}^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{u}^T}_{N \text{ terms}}$$

Symmetric version of previous algorithm gives

$$\begin{aligned} \mathbf{v} &= \sum_{k_2} \cdots \sum_{k_N} \mathbf{u}_{k_2}^{(i)} \cdots \mathbf{u}_{k_N}^{(i)} \mathcal{S}_{k_1, k_2, \ldots, k_N} \\ \mathbf{u}^{(i+1)} &= \frac{\mathbf{v}}{\|\mathbf{v}\|} \end{aligned}$$

Sometimes converges, sometimes not ...

Relax unit norm constraint to unit ball:

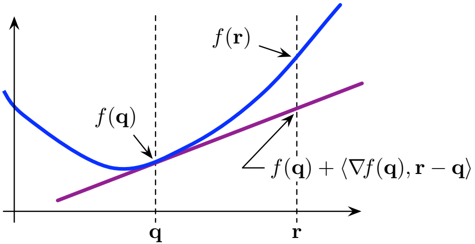
$$\mathcal{B} = \left\{\mathbf{u} : \|\mathbf{u}\| \leq 1\right\}.$$

Special case: $f(\mathbf{u}) = \mathbf{u}^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{u}^T$ is convex (or concave) over $\mathcal{B}$.

Gradient inequality:

$f(\mathbf{r}) \geq \mathbf{f}(\mathbf{q}) + \langle \mathbf{r} - \mathbf{q}, \nabla f(\mathbf{q}) \rangle$

for all $\mathbf{q}, \mathbf{r} \in \mathcal{B}$.



This applies, therefore, to successive iterates: $\mathbf{q} = \mathbf{u}^{(i)}$ and $\mathbf{r} = \mathbf{u}^{(i+1)}$.

From gradient inequality,

$$f(\mathbf{u}^{(i+1)}) - f(\mathbf{u}^{(i)}) \geq \langle \mathbf{u}^{(i+1)} - \mathbf{u}^{(i)}, \nabla f(\mathbf{u}^{(i)}) \rangle$$

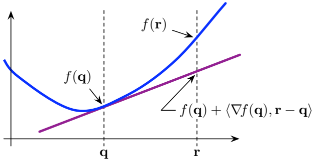
and since $\mathbf{u}^{(i+1)} = \nabla f / \|\nabla f\|$, we have

$$\langle \mathbf{u}^{(i+1)}, \nabla f(\mathbf{u}^{(i)}) \rangle = \|\nabla f(\mathbf{u}^{(i)})\| > \langle \mathbf{u}^{(i)}, \nabla f(\mathbf{u}^{(i)}) \rangle, \qquad \text{if } \mathbf{u}^{(i+1)} \neq \mathbf{u}^{(i)},$$

so that $f(\mathbf{u}^{(i+1)}) > f(\mathbf{u}^{(i)})$.

Remark: The discrepancy

$$D_f(\mathbf{r}, \mathbf{q}) = f(\mathbf{r}) - f(\mathbf{q}) - \langle \mathbf{r} - \mathbf{q}, \nabla f(\mathbf{q}) \rangle \geq 0$$

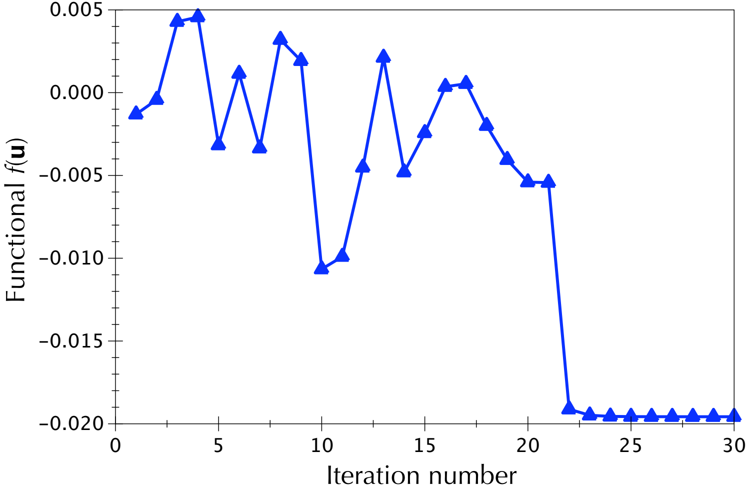


is the Bregman distance induced by f. If f is strongly convex, i.e.,

$$D_f(\mathbf{r}, \mathbf{q}) \geq \alpha \|\mathbf{r} - \mathbf{q}\|^2, \qquad \text{for all } \mathbf{q}, \mathbf{r} \in \mathcal{B},$$

then local convergence rate is at least linear.

What if $f(\mathbf{u})$ is not convex over $\mathcal{B}$?



Gradient interpretation

Rewrite the functional as a "Rayleigh quotient" over $\mathcal{B}$:

$$J(\mathbf{u}) = \frac{f(\mathbf{u})}{\|\mathbf{u}\|^N}, \qquad \mathbf{u} \in \mathcal{B}\backslash\mathbf{0}$$

and consider a gradient ascent algorithm:

$$\begin{aligned} \mathbf{v} &= \mathbf{u}^{(i)} + \mu_i \nabla J(\mathbf{u}^{(i)}) \\ &= \mathbf{u}^{(i)} + \mu_i [\nabla f(\mathbf{u}^{(i)}) - \beta_i \mathbf{u}^{(i)}] \\ \mathbf{u}^{(i+1)} &= \frac{\mathbf{v}}{\|\mathbf{v}\|} \end{aligned}$$

with $\beta_i = \langle \nabla f(\mathbf{u}^{(i)}), \mathbf{u}^{(i)} \rangle$.

Previous algorithm is obtained using $\mu_i = 1/\beta_i$.

Gradient interpretation

Rewrite the functional as a "Rayleigh quotient" over $\mathcal{B}$:

$$J(\mathbf{u}) = \frac{f(\mathbf{u})}{\|\mathbf{u}\|^N}, \qquad \mathbf{u} \in \mathcal{B}\backslash\mathbf{0}$$

and consider a gradient ascent algorithm:

$$\begin{aligned} \mathbf{v} &= \mathbf{u}^{(i)} + \mu_i \nabla J(\mathbf{u}^{(i)}) \\ &= \mathbf{u}^{(i)} + \mu_i [\nabla f(\mathbf{u}^{(i)}) - \beta_i \mathbf{u}^{(i)}] \\ \mathbf{u}^{(i+1)} &= \frac{\mathbf{v}}{\|\mathbf{v}\|} \end{aligned}$$

with $\beta_i = \langle \nabla f(\mathbf{u}^{(i)}), \mathbf{u}^{(i)} \rangle$.

Previous algorithm is obtained using $\mu_i = 1/\beta_i$.

A simple "trick"

Consider "regularized" functional

$$g(\mathbf{u}) = f(\mathbf{u}) + \frac{\gamma}{2}\left(\mathbf{u}^T\mathbf{u} - 1\right), \qquad \mathbf{u} \in \mathcal{B},$$

with γ a free parameter, so that $g(\mathbf{u}) = f(\mathbf{u})$ for $\mathbf{u} \in \partial\mathcal{B}$.

Now,

$$g(\mathbf{u}) \text{ is convex over } \mathcal{B} \quad \Leftrightarrow \quad \nabla^2 g(\mathbf{u}) \geq 0 \text{ over } \mathcal{B}.$$

The Hessians are related as

$$\nabla^2 g(\mathbf{u}) = \nabla^2 f(\mathbf{u}) + \gamma\mathbf{I}$$

so with

$$\lambda_* \triangleq \min_{\mathbf{u}\in\mathcal{B}} \left[\lambda_{\min}\left(\nabla^2 f(\mathbf{u})\right)\right]$$

choose $\gamma \geq -\lambda_*$.

Gradient ascent algorithm for "regularized" functional is no different:

$$\begin{aligned} \mathbf{v} &= \mathbf{u}^{(i)} + \mu_i \left[\nabla g_i - \alpha_i \mathbf{u}^{(i)}\right] \qquad \alpha_i = \langle \nabla g_i, \mathbf{u}^{(i)} \rangle \\ &= \mathbf{u}^{(i)} + \mu_i \left[\nabla f_i + \gamma \mathbf{u}^{(i)} - \beta_i \mathbf{u}^{(i)} - \gamma \mathbf{u}^{(i)}\right] \\ &= \mathbf{u}^{(i)} + \mu_i \left[\nabla f_i - \beta_i \mathbf{u}^{(i)}\right] \\ \mathbf{u}^{(i+1)} &= \frac{\mathbf{v}}{\|\mathbf{v}\|} \end{aligned}$$

This gives $f(\mathbf{u}^{(i+1)}) > f(\mathbf{u}^{(i)})$ if $\mu_i = 1/\alpha_i = 1/(\beta_i + \gamma)$.

Theorem (R. & Kofidis, 2003)

If $\mathbf{u}^{(i)}$ is not a stationary point, then $f(\mathbf{u}^{(i+1)}) > f(\mathbf{u}^{(i)})$ whenever

$$0 < \mu_i < \frac{2(\beta_i - \lambda_*)}{\beta_i^2 + (\beta_i - \lambda_*)^2 - \|\nabla f(\mathbf{u}^{(i)})\|^2}$$

Gradient ascent algorithm for "regularized" functional is no different:

$$\begin{aligned}
\mathbf{v} &= \mathbf{u}^{(i)} + \mu_i \left[\nabla g_i - \alpha_i \mathbf{u}^{(i)}\right] \qquad\qquad \alpha_i = \langle \nabla g_i, \mathbf{u}^{(i)} \rangle \\
&= \mathbf{u}^{(i)} + \mu_i \left[\nabla f_i + \gamma\, \mathbf{u}^{(i)} - \beta_i \mathbf{u}^{(i)} - \gamma\, \mathbf{u}^{(i)}\right] \\
&= \mathbf{u}^{(i)} + \mu_i \left[\nabla f_i - \beta_i \mathbf{u}^{(i)}\right] \\
\mathbf{u}^{(i+1)} &= \frac{\mathbf{v}}{\|\mathbf{v}\|}
\end{aligned}$$

This gives $f(\mathbf{u}^{(i+1)}) > f(\mathbf{u}^{(i)})$ if $\mu_i = 1/\alpha_i = 1/(\beta_i + \gamma)$.

Theorem (R. & Kofidis, 2003)

If $\mathbf{u}^{(i)}$ is not a stationary point, then $f(\mathbf{u}^{(i+1)}) > f(\mathbf{u}^{(i)})$ whenever

$$0 < \mu_i < \frac{2(\beta_i - \lambda_*)}{\beta_i^2 + (\beta_i - \lambda_*)^2 - \|\nabla f(\mathbf{u}^{(i)})\|^2}$$

PARAFAC Decomposition

When a (symmetric or not) tensor *can* be Tucker factored as

$$\mathcal{S} = \mathbf{A}^{(1)} \star \mathbf{A}^{(2)} \star \cdots \star \mathbf{A}^{(N)}$$

uniqueness is addressed by Kruskal ($N = 3$; 1977) and Bro & Sidiropoulos ($N > 3$; 1998), using "k-rank" concepts. (Not always "robust").

Rank-r Approximation: Find rank r matrices $\mathbf{U}^{(i)}$, with $[\mathbf{U}^{(i)}]^T\mathbf{U}^{(i)} = \mathbf{I}_r$, and "core" tensor $\mathcal{T}$ (of size $r \times r \times \cdots \times r$), to minimize Frobenius norm of

$$\mathcal{S} - \mathbf{U}^{(1)} \overset{\mathcal{T}}{\star} \mathbf{U}^{(2)} \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \mathbf{U}^{(N)}$$

- In matrix case, same as r applications of rank-one problem.
- In tensor case, no longer true, unless $\mathcal{T}$ happens to be diagonal.
- In particular, not same as "truncated tensor SVD".

PARAFAC Decomposition

When a (symmetric or not) tensor *can* be Tucker factored as

$$\mathcal{S} = \mathbf{A}^{(1)} \star \mathbf{A}^{(2)} \star \cdots \star \mathbf{A}^{(N)}$$

uniqueness is addressed by Kruskal ($N = 3$; 1977) and Bro & Sidiropoulos ($N > 3$; 1998), using "k-rank" concepts. (Not always "robust").

Rank-r Approximation: Find rank r matrices $\mathbf{U}^{(i)}$, with $[\mathbf{U}^{(i)}]^T\mathbf{U}^{(i)} = \mathbf{I}_r$, and "core" tensor $\mathcal{T}$ (of size $r \times r \times \cdots \times r$), to minimize Frobenius norm of

$$\mathcal{S} - \mathbf{U}^{(1)} \overset{\mathcal{T}}{\star} \mathbf{U}^{(2)} \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \mathbf{U}^{(N)}$$

- In matrix case, same as r applications of rank-one problem.
- In tensor case, no longer true, unless $\mathcal{T}$ happens to be diagonal.
- In particular, not same as "truncated tensor SVD".

PARAFAC Decomposition

When a (symmetric or not) tensor *can* be Tucker factored as

$$\mathcal{S} = \mathbf{A}^{(1)} \star \mathbf{A}^{(2)} \star \cdots \star \mathbf{A}^{(N)}$$

uniqueness is addressed by Kruskal ($N = 3$; 1977) and Bro & Sidiropoulos ($N > 3$; 1998), using "k-rank" concepts. (Not always "robust").

Rank-r Approximation: Find rank r matrices $\mathbf{U}^{(i)}$, with $[\mathbf{U}^{(i)}]^T\mathbf{U}^{(i)} = \mathbf{I}_r$, and "core" tensor $\mathcal{T}$ (of size $r \times r \times \cdots \times r$), to minimize Frobenius norm of

$$\mathcal{S} - \mathbf{U}^{(1)} \overset{\mathcal{T}}{\star} \mathbf{U}^{(2)} \overset{\mathcal{T}}{\star} \cdots \overset{\mathcal{T}}{\star} \mathbf{U}^{(N)}$$

- In matrix case, same as r applications of rank-one problem.
- In tensor case, no longer true, unless $\mathcal{T}$ happens to be diagonal.
- In particular, not same as "truncated tensor SVD".

Symmetric Rank-r Approximation

When $\mathcal{S}$ is symmetric, of dimensions $d \times d \times \cdots d$, introduce tensor

$$\mathcal{R}^{(i)} = \mathbf{I} \overset{\mathcal{S}}{\star} \underbrace{[\mathbf{U}^{(i)}]^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} [\mathbf{U}^{(i)}]^T}_{N-1 \text{ terms}} \qquad (\text{size } d \times r \times \cdots \times r)$$

Its unfolded matrix is (of size $d \times r^{N-1}$)

$$\mathbf{R}^{(i)}_{j,k} = \mathcal{R}^{(i)}_{j,k_2,\ldots,k_N} \qquad \text{with} \qquad k = k_2\, r + k_3\, r^2 + \cdots + k_N\, r^{N-1}$$

Then choose for $\mathbf{U}^{(i+1)}$ (size $d \times r$) the r dominant left singular vectors of $\mathbf{R}^{(i)}$, and iterate.

Convergence

With

$$\mathcal{T} = \underbrace{\mathbf{U}^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{U}^T}_{N \text{ terms}}$$

Can show that

$$\mathcal{T}^{(i+1)} > \mathcal{T}^{(i)}$$

in the sense that

$$\mathbf{v}^T \overset{\mathcal{T}^{(i+1)}}{\star} \cdots \overset{\mathcal{T}^{(i+1)}}{\star} \mathbf{v}^T > \mathbf{v}^T \overset{\mathcal{T}^{(i)}}{\star} \cdots \overset{\mathcal{T}^{(i)}}{\star} \mathbf{v}^T, \quad \forall \mathbf{v} \neq \mathbf{0}$$

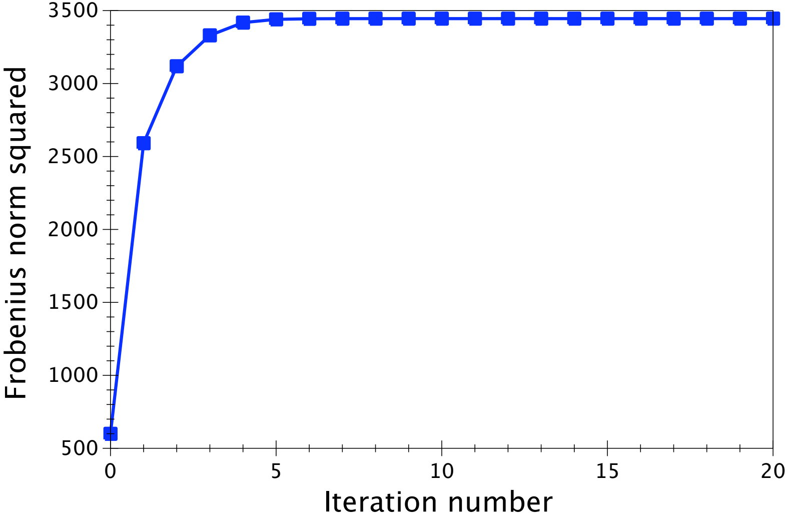
provided

$$f(\mathbf{u}) = \mathbf{u}^T \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{u}^T$$

is convex for $\mathbf{u} \in \mathcal{B}$.

Typical simulation example

$N = 4$, $d = 11$, $r = 5$:



A stationary stochastic process: $r_{|k-l|} = E = (x_{i-k}\, x_{i-l}) = E(x_i\, x_{i+k-l})$ can be modeled as a *linear process*, i.e., $x_i = \sum_j h_j\, \epsilon_{i-j}$ with $\{\epsilon_i\}$ i.i.d, iff

$$\mathbf{R}_N = \begin{bmatrix} r_0 & r_1 & \cdots & r_N \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_N & \cdots & r_1 & r_0 \end{bmatrix} > 0, \quad \text{for all } N$$

Displacement structure:

$$\mathbf{R}_N - \mathbf{Z}\mathbf{R}_N\mathbf{Z}^T = \begin{bmatrix} r_0 & r_1 & \cdots & r_N \\ r_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ r_N & 0 & \cdots & 0 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \sqrt{r_0} \\ r_1/\sqrt{r_0} \\ \vdots \\ r_N\sqrt{r_0} \end{bmatrix}}_{\mathbf{a}} [\cdot]^T - \underbrace{\begin{bmatrix} 0 \\ r_1/\sqrt{r_0} \\ \vdots \\ r_N\sqrt{r_0} \end{bmatrix}}_{\mathbf{b}} [\cdot]^T$$

using shift matrix

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & \cdots \\ 1 & 0 & \ddots & \ddots \\ 0 & 1 & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

Generator formulation: With z_1 and z_2 two complex variables

$$[1 \;\; z_1 \;\; z_1^2 \;\; \cdots] \mathbf{R}_\infty \begin{bmatrix} 1 \\ z_2 \\ z_2^2 \\ \vdots \end{bmatrix} = \frac{a(z_1)\, a(z_2) - b(z_1)\, b(z_2)}{1 - z_1\, z_2}$$

and $\mathbf{R}_\infty \geq 0$ if and only if there exists a Schur function $S(z)$ for which

$$b(z) = S(z)\, a(z)$$

Interpolation problem: Given n points $\{z_i\}$ in unit disk, find $S(z)$ such that

$$b(z_i) = S(z_i)\, a(z_i), \qquad i = 1, 2, \ldots, n.$$

Underlies spectral factorization:

$$H(z) = \frac{z\, Q(z)}{1 - z\, S(z)}, \qquad \text{with} \quad Q(z)\, Q(z^{-1}) + S(z)\, S(z^{-1}) = 1,$$

gives linear model $H(z) = \sum_{k \geq 1} h_k\, z^k$ which generates the correlation data.

Generator formulation: With z_1 and z_2 two complex variables

$$[1 \;\; z_1 \;\; z_1^2 \;\; \cdots\,]\mathbf{R}_\infty \begin{bmatrix} 1 \\ z_2 \\ z_2^2 \\ \vdots \end{bmatrix} = \frac{a(z_1)\,a(z_2) - b(z_1)\,b(z_2)}{1 - z_1\,z_2}$$

and $\mathbf{R}_\infty \geq 0$ if and only if there exists a Schur function $S(z)$ for which

$$b(z) = S(z)\,a(z)$$

Interpolation problem: Given n points $\{z_i\}$ in unit disk, find $S(z)$ such that

$$b(z_i) = S(z_i)\,a(z_i), \qquad i = 1, 2, \ldots, n.$$

Underlies spectral factorization:

$$H(z) = \frac{z\,Q(z)}{1 - z\,S(z)}, \qquad \text{with} \quad Q(z)\,Q(z^{-1}) + S(z)\,S(z^{-1}) = 1,$$

gives linear model $H(z) = \sum_{k \geq 1} h_k\, z^k$ which generates the correlation data.

Consider higher order tensor

$$\mathcal{R}_{j_1, j_2 \ldots, j_N} = \text{cum}_N(x_{i-j_1}, \ldots, x_{i-j_N})$$

which is "Toeplitz" whenever $\{x_i\}$ is stationary. Process is "linear" if we can find factorization

$$\mathcal{R} = \gamma_N \underbrace{\mathbf{H} \star \mathbf{H} \star \cdots \star \mathbf{H}}_{N \text{ terms}} \qquad \text{with} \quad \mathbf{H} = \begin{bmatrix} 0 & h_1 & h_2 & h_3 & \cdots \\ 0 & 0 & h_1 & h_2 & \ddots \\ 0 & 0 & 0 & h_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

Linear or not, the displacement structure

$$\mathcal{S} \triangleq \mathcal{R} - \mathbf{Z} \overset{\mathcal{R}}{\star} \mathbf{Z} \overset{\mathcal{R}}{\star} \cdots \overset{\mathcal{R}}{\star} \mathbf{Z}$$

vanishes except along its faces.

Consider Hankel matrix

$$\mathbf{\Gamma} = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \iddots \\ h_3 & h_4 & h_5 & \iddots \\ \vdots & \iddots & \iddots & \iddots \end{bmatrix} \qquad \Leftrightarrow \qquad H(z) = \sum_{k=1}^{\infty} h_k \, z^k$$

and the identity

$$\mathcal{G} \triangleq \underbrace{\mathbf{\Gamma} \star \cdots \star \mathbf{\Gamma}}_{N \text{ terms}} = \underbrace{\mathbf{H} \star \cdots \star \mathbf{H}}_{\mathcal{R}/\gamma_N} - \mathbf{H}^T \star \cdots \star \mathbf{H}^T$$

$\mathcal{G}$ and $\mathcal{R}/\gamma_N$ agree on faces, and:

$$\mathcal{G} - \mathbf{Z} \overset{\mathcal{G}}{\star} \cdots \overset{\mathcal{G}}{\star} \mathbf{Z} = \mathcal{S}/\gamma_N - \mathbf{h} \star \cdots \star \mathbf{h}$$

with $\mathbf{h} = [0 \;\; h_1 \;\; h_2 \;\; h_3 \;\; \cdots]^T$.

Let $\mathbf{z}_i = [1 \;\; z_i \;\; z_i^2 \;\; z_i^3 \;\; \cdots]$, and introduce generator function

$$S(z_1, z_2, \ldots, z_N) \triangleq \mathbf{z}_1 \overset{\mathcal{S}}{\star} \mathbf{z}_2 \overset{\mathcal{S}}{\star} \cdots \overset{\mathcal{S}}{\star} \mathbf{z}_N$$

Setting $z_N = 1/(z_1 \cdots z_{N-1})$ gives the polyspectrum:

$$\begin{aligned} &S(z_1, \ldots, z_{N-1}, (z_1 \cdots z_{N-1})^{-1}) \\ &= \sum_{k_1, \ldots, k_{N-1}} \mathrm{cum}_N(x_n, x_{n-k_1}, \ldots, x_{n-k_{N-1}})\, z_1^{k_1} \cdots z_{N-1}^{k_{N-1}} \quad \text{(general case)} \\ &= \gamma_N\, H(z_1) \cdots H(z^{N-1})\, H\Big((z_1 \cdots z_{N-1})^{-1}\Big) \quad \text{(when linear)} \end{aligned}$$

Open problem: Given specific evaluation points $\{\alpha_i\}$ in unit disk, when can we find an $H(z)$ whose output cumulants replicate the values[1]

$$S(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_N}) \quad ?$$

[1] V. S. Grigorascu and P. R., "Tensor displacement structures…" in *Fast Algorithms for Matrices w/ Structure*, T. Kailath & A Sayed, eds., SIAM, 1999

Concluding remarks

- Rank one tensor approximation follows easily from matrix case, *except* that it need not lead to a "fundamental" decomposition.
- Symmetric version is convergent if a relaxed function is convex or concave.
- If not, reduce step size (bound obtained analytically).
- Higher-order factorization/interpolation not always so well defined, and offers open terrain.