THE HIGHER-ORDER POWER METHOD REVISITED: CONVERGENCE PROOFS AND EFFECTIVE INITIALIZATION

Phillip A. Regalia and Eleftherios Kofidis[†]

Institut National des Télécommunications
9, rue Charles Fourier
F-91011 Evry cedex, France
`[regalia,kofidis]@sim.int-evry.fr`

ABSTRACT

We revisit the higher-order power method of De Lathauwer *et al.* [1] for rank-one tensor approximation, and its relation to contrast maximization as used in blind deconvolution. We establish a simple convergence proof for the general nonsymmetric tensor case. We show also that a symmetric version of the algorithm, offering an order of magnitude reduction in computational complexity but discarded in [1] as unpredictable, is likewise provably convergent. A new initialization scheme is also developed which, unlike the TSVD-based initialization, leads to a quantifiable proximity to the globally optimal solution.

1. PROBLEM STRUCTURE

Let $\mathcal{T}$ be some k^{th} order real tensor, i.e., a k-index array containing elements $\mathcal{T}_{i_1,i_2,\ldots,i_k}$. If $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$, …, $\mathbf{u}^{(k)}$ are k column vectors, their k^{th}-order "outer product" yields a "rank-one" tensor according to the formula [2]

$$\left[\mathbf{u}^{(1)}\star\mathbf{u}^{(2)}\star\cdots\star\mathbf{u}^{(k)}\right]_{i_1,i_2,\ldots,i_k}=\mathbf{u}^{(1)}_{i_1}\mathbf{u}^{(2)}_{i_2}\cdots\mathbf{u}^{(k)}_{i_k}$$

(For $k=2$ terms, this gives $\mathbf{u}^{(1)}\star\mathbf{u}^{(2)}=\mathbf{u}^{(1)}[\mathbf{u}^{(2)}]^t$). The Frobenius norm of any tensor $\mathcal{T}$ is the square root of the sum of squares of all its entries:

$$\|\mathcal{T}\|_F\triangleq\Big(\sum_{i_1,i_2,\ldots,i_k}\mathcal{T}^2_{i_1,i_2,\ldots,i_k}\Big)^{1/2}.$$

For vectors, the Frobenius and Euclidean (or ℓ_2) norms coincide. The "rank-one tensor approximation" problem posed in [1] is:

Problem 1 *Given a k^{th}-order tensor $\mathcal{T}$, find k unit-norm (column) vectors $\mathbf{u}^{(1)}$, …, $\mathbf{u}^{(k)}$ and a scalar λ which minimize the Frobenius norm squared*

$$\|\mathcal{T}-\lambda[\mathbf{u}^{(1)}\star\cdots\star\mathbf{u}^{(k)}]\|_F^2=\sum_{i_1,\ldots,i_k}\big(\mathcal{T}_{i_1,\ldots,i_k}-\lambda\,\mathbf{u}^{(1)}_{i_1}\cdots\mathbf{u}^{(k)}_{i_k}\big)^2.$$

To characterize solutions, let vec$(\cdot)$ rearrange a tensor as a vector according to

$$\mathrm{vec}(\mathcal{T})=\big[\mathcal{T}_{0,\ldots,0,0}\;\mathcal{T}_{0,\ldots,0,1}\;\cdots\;\mathcal{T}_{0,\ldots,1,0}\;\mathcal{T}_{0,\ldots,1,1}\;\cdots\big]^t.$$

Since the Frobenius norm is invariant to a rearrangement of its tensor argument, we see that

$$\|\mathcal{T}-\lambda[\mathbf{u}^{(1)}\star\cdots\star\mathbf{u}^{(k)}]\|_F=\|\underbrace{\mathrm{vec}(\mathcal{T})}_{\triangleq\,\mathbf{a}}-\lambda\,\underbrace{\mathrm{vec}(\mathbf{u}^{(1)}\star\cdots\star\mathbf{u}^{(k)})}_{\triangleq\,\mathbf{b}}\|_F$$

If we fix the $\mathbf{u}^{(i)}$, this becomes a standard least-squares problem in λ, whose optimal solution is given by

$$\begin{aligned}\lambda_{opt}&=\frac{\mathbf{a}^t\mathbf{b}}{\mathbf{b}^t\mathbf{b}}=\mathbf{a}^t\mathbf{b}\quad\text{since }\mathbf{b}^t\mathbf{b}=1\\&=\sum_{i_1,\ldots,i_k}\mathcal{T}_{i_1,\ldots,i_k}\mathbf{u}^{(1)}_{i_1}\cdots\mathbf{u}^{(k)}_{i_k}\triangleq f(\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(k)})\end{aligned}$$

in which $\mathbf{b}^t\mathbf{b}=1$ because the vectors $\mathbf{u}^{(i)}$ all have unit norm. As this optimal choice of λ renders the error $\mathbf{a}-\lambda_{opt}\mathbf{b}$ orthogonal to the approximant $\lambda_{opt}\mathbf{b}$, we get

$$\begin{aligned}\|\mathbf{a}-\lambda_{opt}\mathbf{b}\|^2&=\|\mathbf{a}\|^2-|\lambda_{opt}|^2\|\mathbf{b}\|^2\\&=\|\mathcal{T}\|_F^2-|f(\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(k)})|^2.\qquad(1)\end{aligned}$$

Any set $\{\mathbf{u}^{(i)}\}$ which attains a local minimum in Problem 1 must therefore attain a local maximum of $|f|$, giving:

Theorem 1 *The k unit norm vectors $\mathbf{u}^{(1)}$, …, $\mathbf{u}^{(k)}$ correspond to a local minimum of Problem 1 if and only if these same vectors yield a local maximum of $|f(\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(k)})|$, where*

$$f(\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(k)})=\sum_{i_1,\ldots,i_k}\mathcal{T}_{i_1,\ldots,i_k}\mathbf{u}^{(1)}_{i_1}\cdots\mathbf{u}^{(k)}_{i_k}.$$

The corresponding value of λ is $\lambda=f(\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(k)})$.

[†] E. Kofidis is supported by a Training and Mobility of Researchers grant of the European Commission, contract no. ERBFMBICT982959.

The stationary points of the functional f have been calculated in [1], leading to the system of equalities

$$\begin{aligned}\sum_{i_2,\ldots,i_k} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_2}^{(2)} \cdots \mathbf{u}_{i_k}^{(k)} &= \lambda \mathbf{u}_{i_1}^{(1)} \\ \sum_{i_1,i_3,\ldots,i_k} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_1}^{(1)} \mathbf{u}_{i_3}^{(3)} \cdots \mathbf{u}_{i_k}^{(k)} &= \lambda \mathbf{u}_{i_2}^{(2)} \\ &\vdots \\ \sum_{i_1,\ldots,i_{k-1}} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_1}^{(1)} \cdots \mathbf{u}_{i_{k-1}}^{(k-1)} &= \lambda \mathbf{u}_{i_k}^{(k)}\end{aligned}$$

[For two-index tensors, i.e., matrices, this system identifies $(|\lambda|, \mathbf{u}^{(1)}, \pm\mathbf{u}^{(2)})$ as a singular triple of the matrix $\mathcal{T}$]. For the k^{th}-order tensor case, an iterative algorithm is proposed in [1] of the form

$$\begin{aligned}\mathbf{u}_{i_1}^{(1)}(n+1) &= \frac{\varepsilon}{\|\cdot\|} \sum_{i_2,\ldots,i_k} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_2}^{(2)}(n) \cdots \mathbf{u}_{i_k}^{(k)}(n) \qquad (2) \\ \mathbf{u}_{i_2}^{(2)}(n+1) &= \frac{\varepsilon}{\|\cdot\|} \sum_{i_1,i_3\ldots,i_k} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_1}^{(1)}(n+1) \mathbf{u}_{i_3}^{(3)}(n) \cdots \mathbf{u}_{i_k}^{(k)}(n) \\ &\vdots \\ \mathbf{u}_{i_k}^{(k)}(n+1) &= \frac{\varepsilon}{\|\cdot\|} \sum_{i_1,\ldots,i_{k-1}} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_1}^{(1)}(n+1) \cdots \mathbf{u}_{i_{k-1}}^{(k-1)}(n+1)\end{aligned}$$

in which $\varepsilon = +1$ or $\varepsilon = -1$; the divisors $\|\cdot\|$ in each expression simply scale each $\mathbf{u}^{(i)}(n+1)$ to unit norm. The fixed points of this iteration clearly satisfy the equations for a stationary point above, but convergence of this procedure is not fully proved in [1]. We offer a proof as:

Theorem 2 *If $\varepsilon = +1$ (resp., $\varepsilon = -1$), the iterative algorithm converges to a local maximum (resp., local minimum) of the functional f, save for an exceptional set of initial conditions on the $\{\mathbf{u}^{(i)}\}$.*[1]

The proof consists in noting that f is a multilinear function of its vector arguments. Isolating its dependence on $\mathbf{u}^{(1)}$, we may write an inner product as

$$f\Big(\mathbf{u}^{(1)}(n), \mathbf{u}^{(2)}(n), \ldots, \mathbf{u}^{(k)}(n)\Big) = \langle \mathbf{u}^{(1)}(n), \mathbf{x} \rangle,$$

using the vector $\mathbf{x}$ given by

$$\mathbf{x}_{i_1} = \sum_{i_2,\ldots,i_k} \mathcal{T}_{i_1,i_2,\ldots,i_k} \mathbf{u}_{i_2}^{(2)}(n) \cdots \mathbf{u}_{i_k}^{(k)}(n).$$

The change in f upon adapting $\mathbf{u}^{(1)}$ in (2) becomes

$$\langle \mathbf{u}^{(1)}(n+1), \mathbf{x} \rangle - \langle \mathbf{u}^{(1)}(n), \mathbf{x} \rangle.$$

Since $\|\mathbf{u}^{(1)}\| = 1$, the Cauchy-Schwarz inequality gives

$$-\|\mathbf{x}\| \leq \langle \mathbf{u}^{(1)}, \mathbf{x} \rangle \leq \|\mathbf{x}\|.$$

The upper (resp., lower) bound is obtained if and only if $\mathbf{u}^{(1)} = \mathbf{x}/\|\mathbf{x}\|$ (resp., $\mathbf{u}^{(1)} = -\mathbf{x}/\|\mathbf{x}\|$), which results for $\mathbf{u}^{(1)}(n+1)$ when $\varepsilon = +1$ (resp., $\varepsilon = -1$). Assuming the $\mathbf{u}^{(i)}$ are not at a stationary point, the choice $\varepsilon = +1$ gives $\langle \mathbf{u}^{(1)}(n+1), \mathbf{x} \rangle - \langle \mathbf{u}^{(1)}(n), \mathbf{x} \rangle > 0$, or

$$\begin{aligned}& f\Big(\mathbf{u}^{(1)}(n+1), \mathbf{u}^{(2)}(n), \ldots, \mathbf{u}^{(k)}(n)\Big) \\ & \quad > f\Big(\mathbf{u}^{(1)}(n), \mathbf{u}^{(2)}(n), \ldots, \mathbf{u}^{(k)}(n)\Big).\end{aligned}$$

Repeating this argument for the remaining vectors leads to the chain of inequalities

$$\begin{aligned}& f\Big(\mathbf{u}^{(1)}(n+1), \mathbf{u}^{(2)}(n+1), \ldots, \mathbf{u}^{(k)}(n)\Big) \\ & \quad > f\Big(\mathbf{u}^{(1)}(n+1), \mathbf{u}^{(2)}(n), \ldots, \mathbf{u}^{(k)}(n)\Big) \\ & \vdots \\ & f\Big(\mathbf{u}^{(1)}(n+1), \ldots, \mathbf{u}^{(k-1)}(n+1), \mathbf{u}^{(k)}(n+1)\Big) \\ & \quad > f\Big(\mathbf{u}^{(1)}(n+1), \ldots, \mathbf{u}^{(k-1)}(n+1), \mathbf{u}^{(k)}(n)\Big)\end{aligned}$$

which shows that f is increasing when $\varepsilon = +1$; a dual argument shows f to be decreasing when $\varepsilon = -1$. Since $|f| \leq \|\mathcal{T}\|_F$ [cf. (1)], we conclude that f converges to a local extremum. ◇

2. SYMMETRIC TENSOR CASE

A tensor $\mathcal{T}$ is symmetric if it is invariant to any permutation of its indices; an optimal rank-one approximant in this case should likewise be symmetric, of the form

$$\lambda [\mathbf{u} \star \mathbf{u} \star \cdots \star \mathbf{u}]_{i_1,i_2,\ldots,i_k} = \lambda \mathbf{u}_{i_1} \mathbf{u}_{i_2} \cdots \mathbf{u}_{i_k}.$$

The symmetric-tensor counterpart to Theorem 1 follows similarly: A unit-norm vector $\mathbf{u}$ and scalar λ minimize the Frobenius norm $\|\mathcal{T} - \lambda\, \mathbf{u} \star \cdots \star \mathbf{u}\|_F$ if and only if this unit-norm $\mathbf{u}$ maximizes $|f(\mathbf{u})|$, with

$$f(\mathbf{u}) = \sum_{i_1,\ldots,i_k} \mathcal{T}_{i_1,\ldots,i_k} \mathbf{u}_{i_1} \cdots \mathbf{u}_{i_k}, \quad \sum_i \mathbf{u}_i^2 = 1, \qquad (3)$$

for which the corresponding value of λ is $\lambda = f(\mathbf{u})$.

Example 1. Let $\mathbf{x} \in \mathbb{R}^p$ be a vector comprised of random variables, and let $\mathcal{T}$ be its k^{th}-order cumulant tensor of dimensions $p \times p \times \cdots \times p$ defined as

$$\mathcal{T}_{i_1,i_2,\ldots,i_k} = \mathrm{cum}[\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_k}].$$

Since cumulants are symmetric functions of their arguments, the tensor $\mathcal{T}$ is symmetric (invariant to any permutation of its indices). Suppose the elements of $\mathbf{x}$ are

---

[1] The exceptional set will include all saddle points and all crest leading to such saddle points.

uncorrelated to second order, i.e., $E[\mathbf{x}_i\mathbf{x}_j] = \delta_{ij}$. If we set $y = \mathbf{u}'\mathbf{x}$, and denote by $\mathrm{cum}_k(y)$ the k^{th}-order cumulant of y, then

$$\frac{f(\mathbf{u})}{\|\mathbf{u}\|^k} = \frac{\mathrm{cum}_k(y)}{[\mathrm{cum}_2(y)]^{k/2}},$$

which, for k even, yields a common contrast function for blind deconvolution [4], [5]. ◇

A natural question (left open in [1]) is whether a symmetric version of the iterative algorithm (2), *viz.*

$$\begin{aligned} \tilde{\mathbf{u}}_{i_1} &= \sum_{i_2,\ldots,i_k} \mathcal{T}_{i_2,\ldots,i_k}\mathbf{u}_{i_2}(n)\cdots\mathbf{u}_{i_k}(n) \\ \mathbf{u}(n+1) &= \tilde{\mathbf{u}}/\|\tilde{\mathbf{u}}\| \end{aligned} \tag{4}$$

remains convergent to an extremum of $|f(\mathbf{u})|$.

To this end, let $\mathbf{w}$ be a free vector in $\mathbb{R}^p$ (i.e., not necessarily of unit norm). The function $f(\mathbf{w})$ is convex if, for all $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ in $\mathbb{R}^p$, and all $0 \leq \alpha \leq 1$,

$$f\big(\alpha\mathbf{w}^{(1)} + (1-\alpha)\mathbf{w}^{(2)}\big) \leq \alpha f(\mathbf{w}^{(1)}) + (1-\alpha)f(\mathbf{w}^{(2)}).$$

The function $f(\mathbf{w})$ is concave if $-f(\mathbf{w})$ is convex.

Example 2. Suppose that the vector $\mathbf{x}$ is a mixture of indepedent sources collected into a column vector $\mathbf{s}$, as in

$$\mathbf{x} = \mathbf{H}\mathbf{s}$$

With appropriate choices of $\mathbf{H}$ and $\mathbf{s}$, this accommodates the noisy convolutional mixture model (e.g., [5]). If we set $\mathbf{g} = \mathbf{w}'\mathbf{H}$, then we have

$$f(\mathbf{w}) = \sum_i \mathbf{g}_i^k \gamma_i$$

where γ_i is the k^{th} order cumulant of the i^{th} component of $\mathbf{s}$. If k is even and all the source cumulants are nonnegative (resp., nonpositive), then $f(\mathbf{w})$ is convex (resp., concave). ◇

Remark: Even if $f(\mathbf{w})$ is convex, its restriction to the set of unit-norm vectors $\mathbf{u} = \mathbf{w}/\|\mathbf{w}\|$ [cf. (3)] is *not* a convex function, since the set of unit-norm vectors is not a convex set.

Theorem 3 *Suppose k is even and that $f(\mathbf{w})$ is a convex (resp., concave) function over $\mathbb{R}^p$. The iterative algorithm (4) then converges to a maximum (resp., minimum) of the function $f(\mathbf{u})/\|\mathbf{u}\|^k$.*

The proof treats the case in which $f(\mathbf{w})$ is convex, as the concave case follows similarly. The (sub-) gradient inequality for convex functions [3] gives

$$f(\mathbf{w}^{(2)}) \geq f(\mathbf{w}^{(1)}) + \langle \mathbf{w}^{(2)} - \mathbf{w}^{(1)}, \nabla f(\mathbf{w}^{(1)})\rangle$$

for all $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ in $\mathbb{R}^p$, where $\nabla f(\mathbf{w})$ is the gradient vector:

$$[\nabla f(\mathbf{w})]_{i_1} = \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}_{i_1}} = k\sum_{i_2,\ldots,i_k} \mathcal{T}_{i_1,i_2,\ldots,i_k}\mathbf{w}_{i_2}\cdots\mathbf{w}_{i_k}.$$

We may therefore set $\mathbf{w}^{(2)} = \mathbf{u}(n+1)$ and $\mathbf{w}^{(1)} = \mathbf{u}(n)$ to obtain

$$\begin{aligned} &f\big(\mathbf{u}(n+1)\big) - f\big(\mathbf{u}(n)\big) \geq \\ &\qquad k\Big(\big\langle \mathbf{u}(n+1), \nabla f\big(\mathbf{u}(n)\big)\big\rangle - \big\langle \mathbf{u}(n), \nabla f\big(\mathbf{u}(n)\big)\big\rangle\Big). \end{aligned}$$

It suffices to show that the right-hand side is positive whenever $\mathbf{u}(n+1) \neq \mathbf{u}(n)$. Now, for any unit-norm $\mathbf{u}$, we have by the Cauchy-Schwarz inequality

$$\big\langle \mathbf{u}, \nabla f\big(\mathbf{u}(n)\big)\big\rangle \leq \big\|\nabla f\big(\mathbf{u}(n)\big)\big\|$$

with equality iff $\mathbf{u} = \nabla f(\mathbf{u}(n))/\|\nabla f(\mathbf{u}(n))\|$. Since this is precisely the formula for $\mathbf{u}(n+1)$, we conclude that

$$\big\langle \mathbf{u}(n+1), \nabla f\big(\mathbf{u}(n)\big)\big\rangle - \big\langle \mathbf{u}(n), \nabla f\big(\mathbf{u}(n)\big)\big\rangle > 0,$$

which implies that $f\big(\mathbf{u}(n)\big)$ is an increasing sequence. ◇

3. INITIALIZATION

The function $f(\mathbf{u})$ has multiple extrema in general, so that convergence to a global extremum is strongly dependent on the initialization. An initialization strategy proposed in [1] uses the TSVD; here we propose an alternate initialization strategy which, in many simulations tested, lies even closer to the global extremum. For simplicity we consider symmetric fourth-order tensors, as the extensions to arbitrary-order nonsymmetric tensors are straightforward.

The Kronecker product $\mathbf{A}\otimes\mathbf{B}$ between two matrices yields a larger matrix containing block elements $a_{ij}\mathbf{B}$. A vec-permutation matrix $\mathbf{P}_v$ is defined [6] from the relation

$$\mathbf{P}_v(\mathbf{A}\otimes\mathbf{B})\mathbf{P}_v' = \mathbf{B}\otimes\mathbf{A},$$

and is dependent only on the dimensions of the matrices $\mathbf{A}$ and $\mathbf{B}$ in question [6].

We now consider a fourth-order $p \times p \times p \times p$ symmetric tensor $\mathcal{T}$. Consider an index mapping from four to two indices, according to

$$\begin{aligned} m &= pi_1 + i_2, \quad 0 \leq i_1, i_2 \leq p-1; \\ n &= pi_3 + i_4, \quad 0 \leq i_3, i_4 \leq p-1. \end{aligned}$$

We may then remap the tensor $\mathcal{T}$ into a matrix $\mathbf{T}$ as

$$[\mathbf{T}]_{m,n} = \mathcal{T}_{i_1,i_2,i_3,i_4}.$$

If $\mathcal{T}$ is symmetric, then the matrix $\mathbf{T}$ satisfies the symmetry relation $\mathbf{T} = \mathbf{T}'$. In fact, a stronger "super-symmetry" relation then results, of the form $\mathbf{P}_v\mathbf{T}\mathbf{P}_v' = \mathbf{T}$, where $\mathbf{P}_v$ is the $p^2 \times p^2$ vec-permutation matrix.

Consider the singular value decomposition of the matrix $\mathbf{T}$:

$$\mathbf{T}\xi_k = \sigma_k\eta_k, \quad \mathbf{T}'\eta_k = \sigma_k\xi_k, \quad k = 1, 2, \ldots, p^2,$$

where the singular values σ_k are nonnegative and the singular vectors ξ_k and η_k form orthonormal bases. From the symmetry property $\mathbf{T} = \mathbf{T}'$, we have $\xi = \pm\eta_k$, while the "super-symmetry" $\mathbf{P}_v\mathbf{T}\mathbf{P}_v = \mathbf{T}$ gives $\xi_k = \mathbf{P}_v'\xi_k$.

We observe now the identity

$$f(\mathbf{u}) = \sum_{i_1,\ldots,i_4} \mathcal{T}_{i_1,\ldots,i_4}\mathbf{u}_{i_1}\cdots\mathbf{u}_{i_4} = (\mathbf{u}\otimes\mathbf{u})'\mathbf{T}(\mathbf{u}\otimes\mathbf{u}).$$

If $\mathbf{u}$ has unit norm, then so does $\mathbf{u}\otimes\mathbf{u}$, so that

$$\left|f(\mathbf{u})\right| = \left|(\mathbf{u}\otimes\mathbf{u})'\mathbf{T}(\mathbf{u}\otimes\mathbf{u})\right| \leq \sigma_1, \quad \text{for all } \|\mathbf{u}\| = 1. \quad (5)$$

Equality results if and only if $\mathbf{u}\otimes\mathbf{u}$ concides with a corresponding singular vector ξ_1 (or $\eta_1 = \pm\xi_1$), but ξ_1 is not a "Kronecker square" in general.

We can nonetheless choose $\mathbf{u}$ to minimize the subspace angle θ between ξ_I and $\mathbf{u}\otimes\mathbf{u}$, as given by

$$\cos\theta = \frac{\xi_1'(\mathbf{u}\otimes\mathbf{u})}{\|\xi_1\|\cdot\|\mathbf{u}\otimes\mathbf{u}\|} = \xi_1'(\mathbf{u}\otimes\mathbf{u})$$

where we note that both ξ_1 and $\mathbf{u}\otimes\mathbf{u}$ have unit norm. We can now observe the identity

$$\xi_1'(\mathbf{u}\otimes\mathbf{u}) = \mathbf{u}'\,\mathrm{unvec}(\xi_1)\mathbf{u}$$

where unvec(ξ_1) returns a matrix by partitioning the column vector ξ_1 into sub-vectors of length p, each of which generates one column of the matrix unvec(ξ_1).

Property 1 *The symmetry relation* $\xi_1 = \mathbf{P}_v'\xi_1$ *is equivalent to symmetry of the matrix* $\mathrm{unvec}(\xi_1) = [\mathrm{unvec}(\xi_1)]'$.

A proof follows readily from the structure of vec-permutation matrices [6] and is omitted. Now,

$$|\xi_1'(\mathbf{u}\otimes\mathbf{u})| = |\mathbf{u}'\,\mathrm{unvec}(\xi_1)\mathbf{u}| \leq \varsigma_1$$

where ς_1 is the largest singular value of unvec(ξ_1); we therefore take $\mathbf{u}$ as the corresponding singular vector to force equality. (Since $\|\mathrm{unvec}(\xi)\|_F = \|\xi\| = 1$, we have $\varsigma_1 \leq 1$).

Returning to the singular value decomposition of $\mathbf{T}$, this choice of initialization for $\mathbf{u}$ leads to

$$f(\mathbf{u}) = (\mathbf{u}\otimes\mathbf{u})'\mathbf{T}(\mathbf{u}\otimes\mathbf{u}) = \varepsilon_1\sigma_1\varsigma_1^2 + \sum_{k=2}^{p^2}\varepsilon_k\sigma_k|\xi_k'(\mathbf{u}\otimes\mathbf{u})|^2$$

since $|\xi_1'(\mathbf{u}\otimes\mathbf{u})| = \varsigma_1$; here ε_k is $+1$ if $\eta_k = \xi_k$, or $\varepsilon_k = -1$ if $\eta_k = -\xi_k$. When $\mathbf{T}$ is sign (semi-) definite (cf. Example 2), the ε_k all take the same sign, so that

$$\left|f(\mathbf{u})\right| = \left|(\mathbf{u}\otimes\mathbf{u})'\mathbf{T}(\mathbf{u}\otimes\mathbf{u})\right| \geq \sigma_1\varsigma_1^2.$$

Combining with the bound (5), we see that when $\varsigma_1 \approx 1$, the $\mathbf{u}$ so initialized must lie near the global maximum of $|f|$. We have observed in numerous simulations that this choice of $\mathbf{u}$ indeed lies in a basin of attraction of the globally optimal unit-norm vector, as does the initialization scheme of [1]. As an example, we consider a 5-input/3-output mixture model using (to three significant figures)

$$\mathbf{H} = \begin{bmatrix} -0.365 & -0.664 & -0.062 & -0.513 & -0.399 \\ 0.365 & 0.156 & -0.599 & -0.626 & 0.303 \\ -0.274 & 0.680 & 0.360 & -0.486 & -0.311 \end{bmatrix}$$

where the five source cumulants are $\gamma_1 = -2$, $\gamma_2 = -1.5$, $\gamma_3 = 1$, $\gamma_4 = 1.5$ and $\gamma_5 = 0$. Using a 3-element vector for $\mathbf{u}$, the function f has one minimum where $f \approx -1.29$ (the global extremum), one maximum where $f \approx 1.19$ (a local extremum), plus saddle points. Figure 1 shows the evolution of the criterion $f[\mathbf{u}(n)]$ for the proposed initialization and that using the TSVD; the new initialization is observed to lie even closer to the optimal value.

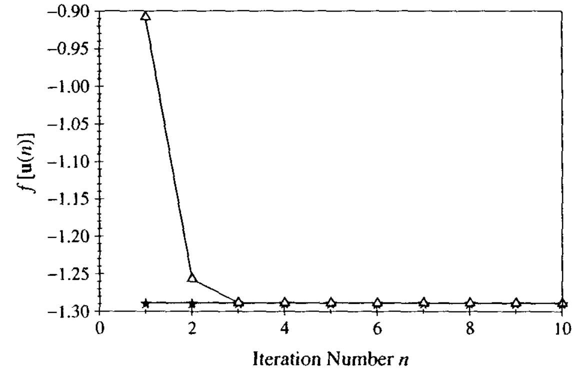

Figure 1: Evolution of $f[\mathbf{u}(n)]$ versus n using the TSVD initialization (△) and the proposed initialization ($\star$).

4. REFERENCES

[1] L. De Lathauwer, P. Comon, B. De Moor, and J. Vandewalle, "Higher-order power method," *Proc. NOLTA-95*, Las Vegas, NV, December 1995.

[2] V. S. Grigorascu and P. A. Regalia, "Tensor displacement structures and polyspectral matching," chap. 9 in: *Fast Reliable Algorithms for Structured Matrices*, T. Kailath and A. H. Sayed, eds., SIAM, Phildelphia, 1999.

[3] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, 1970.

[4] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of non-minimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, pp. 312–321, March 1990.

[5] P. A. Regalia and M. Mboup, "Undermodeled equalization of noisy muli-user channels," *IEEE Wkshp. Nonlinear Signal Image Proc.*, Antalya, Turkey, June 1999.

[6] H. V. Henderson and S. R. Searle, "The vec-permutation matrix, the vec operator and Kronecker products: A review," *Linear and Multilinear Algebra*, vol. 9, pp. 271–288, 1981.