Optimal Joint Azimuth-Elevation and Signal-Array Response Estimation Using Parallel Factor Analysis

R. Bro
Chemometrics Group, Food Tech.
The Royal Vet. & Agri. Univ.
Rolighedsvej 30, DK-1958
Frederiksberg C, Denmark
rb@kvl.dk

N. D. Sidiropoulos and G. B. Giannakis*
Dept. of Electrical Engineering
University of Virginia
Charlottesville VA 22903 - U.S.A.
nikos@virginia.edu
georgios@virginia.edu

Abstract

We consider deterministic joint azimuth-elevation, signal, and array response estimation, and establish a direct link to parallel factor (PARAFAC) analysis, a tool with roots in linear algebra for multi-way arrays. This link affords a powerful identifiability result, plus the opportunity to tap on and extend the available expertise for fitting the PARAFAC model, to derive a deterministic (least squares) joint estimation algorithm, also applicable to multiple-parameter / multiple-invariance ESPRIT subspace fitting problems. These and other issues are demonstrated in pertinent simulation experiments.

1. Data Model

In joint azimuth-elevation estimation (e.g., [1, 2, 3, 4]) using an uncalibrated array of P identical displaced sensor K-tuples (an extension of the usual ESPRIT scenario), M narrowband sources, and a collection of N snapshots from each sensor, the baseband-equivalent model for the array outputs can be written as follows:

$$\begin{aligned}\mathbf{X}_0 &= \mathbf{A}\mathbf{\Phi}_0\mathbf{S} + \mathbf{V}_0\\ \mathbf{X}_1 &= \mathbf{A}\mathbf{\Phi}_1\mathbf{S} + \mathbf{V}_1\\ &\vdots\\ \mathbf{X}_{P-1} &= \mathbf{A}\mathbf{\Phi}_{P-1}\mathbf{S} + \mathbf{V}_{P-1}\end{aligned}$$

where $(\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_{P-1})$ denotes the data, $\mathbf{A}$ is the $K \times M$ (sub)array gain matrix, $\mathbf{S}$ is an $M \times N$ signal matrix, $\mathbf{\Phi}_0, \mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_{P-1}$, are $M \times M$ diagonal matrices containing azimuth and elevation information, and $\mathbf{V}_0, \mathbf{V}_1, \cdots, \mathbf{V}_{P-1}$ model measurement noise.

*Supported by NSF/CAREER CCR-9733540

2. Parallel Factor Analysis

We introduce some notation that will be useful in the sequel. Let $\mathbf{H}$ be the $P \times M$ matrix whose $(p+1)$-th row is the diagonal of $\mathbf{\Phi}_p$, and let $\mathbf{D}_p(\mathbf{H})$ denote the diagonal matrix containing the $(p+1)$-th row of $\mathbf{H}$: $\mathbf{D}_p(\mathbf{H}) = \mathbf{\Phi}_p$, and

$$\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S} + \mathbf{V}_p,\ p = 0, \cdots, P-1.$$

Temporarily ignoring the noise term, and letting $x_p(k,n)$ stand for the (k,n) element of $\mathbf{X}_p$ (output of the k-th element of the p-th subarray at time n), we have:

$$x_p(k,n) = \sum_{m=0}^{M-1} a(k,m)h(p,m)s(m,n). \qquad (1)$$

The model above is well-known in Chemometrics[1] [11], where it is used for spectrophotometric, chromatographic, and flow injection analyses; it is commonly referred to as the PARAFAC (PARAllel FACtor analysis) model [10], or as the *trilinear* model [8, 9]. A distinguishing feature of the trilinear model is its uniqueness: under mild conditions (quite unlike the unconstrained bilinear model), the trilinear model does not suffer from rotational freedom (i.e., $\mathbf{H}, \mathbf{S}, \mathbf{A}$ are identifiable without unitary matrix ambiguities). Several results regarding uniqueness of this model have been known since the 70's. Among them, the following (due to J. Kruskal [8]) is probably the deepest:

[1]Multi-way analysis has also been used in the context of higher-order statistics (under source-independence assumptions) [12, 13]; however, we (in [14, 15]) take a deterministic (least squares) approach in linking multi-way analysis tools with signal processing and communications problems.

Theorem 1 *Define a* $K \times N \times P$ three-way array $\underline{\mathbf{X}}$*, with typical element* $x_{k,n,p} := x_p(k,n)$*. For real-valued arrays,* $\mathbf{S}$, $\mathbf{H}$*, and* $\mathbf{A}$ *can be uniquely identified in the noiseless case (up to the inherently unresolvable permutation and scale ambiguity) from* $\underline{\mathbf{X}}$*, provided that*

$$k_{\mathbf{A}} + k_{\mathbf{H}} + k_{\mathbf{S}^T} \geq 2(M+1), \qquad (2)$$

where $k_{\mathbf{A}}$ *stands for the k*-rank *of* $\mathbf{A}$*: a matrix has k-rank* ℓ *if* every ℓ *(but not* every $\ell+1$*) of its columns are linearly independent (k-rank* $\leq$ *rank, with equality achieved if the matrix is full column rank). The result holds for* $M = 1$ *irrespective of condition (2), so long as* $\underline{\mathbf{X}}$ *does not contain an identically zero 2-D slice along any dimension.*

With real modulations, the elements of all arrays involved are real numbers, and the result applies directly. For complex modulations, the elements of all arrays involved are drawn from the complex field, and, under full-rank conditions, a simpler proof can be constructed using generalized eigenanalysis. Interestingly, it turns out that the base field from which the array elements are drawn makes a difference for such basic concepts as *rank* of multi-way arrays, and complex modulations require a generalization of the line of argument in [8] (see also [14, 15]). Note that condition (2) is *sufficient*: even though it is, in a sense, the deepest result known so far regarding uniqueness of the PARAFAC model, it is not necessary.

The perfect symmetry of the trilinear model in (1) allows two revealing data rearrangements, which can be interpreted as "slicing" the three-way array $\underline{\mathbf{X}}$ along different dimensions. In particular,

$$\mathbf{Y}_k = \mathbf{S}^T \mathbf{D}_k(\mathbf{A}) \mathbf{H}^T, \quad k = 0, 1, \cdots, K-1, \qquad (3)$$

where the $N \times P$ matrix $\mathbf{Y}_k := [x_{k,\cdot,\cdot}]$. Similarly,

$$\mathbf{Z}_n = \mathbf{H} \mathbf{D}_n(\mathbf{S}^T) \mathbf{A}^T, \quad n = 0, 1, \cdots, N-1, \qquad (4)$$

where the $P \times K$ matrix $\mathbf{Z}_n := [q_{\cdot,n,\cdot}]$. These alternative "views" of the data are useful in understanding the core of the algorithm described briefly below (and in detail in [15]). Also note the following corollaries, which can also be obtained simply by data rearrangement and Generalized Eigenanalysis similar to ESPRIT: (i) if $\mathbf{S}$ is fat ($N \geq M$) and full-rank (persistence of excitation), and the DF matrix $\mathbf{H}$ is tall ($P \geq M$) and full-rank, then $K = 2$ suffices for identifiability (provided one of the two diagonals in (3) is non-singular and the two are non-proportional) $\Rightarrow$ more sources than sensors per subarray possible; and (ii) Similarly, if $\mathbf{H}$ is tall and full-rank, the mixing matrix $\mathbf{A}$ is tall ($K \geq M$) and full-rank, and under similar conditions for the diagonals in (4), then $N = 2$ suffices for identifiability $\Rightarrow$ can accommodate very fast variation in fading / azimuth-elevation parameters. Finally, if all matrices involved are full k-rank, then Condition (2) reduces to:

$$\min(K, M) + \min(N, M) + \min(P, M) \geq 2M + 2.$$

2.1. Trilinear Alternating Least Squares Regression

The well-known principle of Alternating Least Squares (ALS) can be used to fit the trilinear model on the basis of noisy observations, possibly initialized using generalized eigenanalysis. This has been proven to be the method of choice in applications of PARAFAC in Chemometrics, for a variety of reasons outlined in [11]. The *trilinear ALS* method is appealing, primarily because it is guaranteed to converge, but also because of its relative simplicity and good performance [11]. Experience shows that, with reasonable initialization, convergence to the global minimum is usually achieved [11]. The explanation (and fundamental difference *vis-a-vis* unstructured bilinear ALS) lies in the inherent *uniqueness* of the trilinear model. Effective estimation of the complex trilinear model involves several sub-algorithms: our fast algorithm (COMFAC) in [15] incorporates compression, initialization using Direct TriLinear Decomposition (DTLD), a compound ALS / Gauss-Newton fit in compressed space, followed by de-compression and a few "guard" ALS steps in uncompressed space. DTLD generates two optimal pseudo-slabs from the available data, and extracts initial estimates using generalized eigenanalysis. Here we confine ourselves to (briefly) explaining how one may utilize the three different ways of slicing the 3-D data to come up with conditional least squares updates for each of the three factors. From the first way of slicing the data, it follows that LS fitting amounts to minimizing:

$$\left\| \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix} - \begin{bmatrix} \mathbf{A}\mathbf{D}_0(\mathbf{H}) \\ \vdots \\ \mathbf{A}\mathbf{D}_{P-1}(\mathbf{H}) \end{bmatrix} \mathbf{S} \right\|_F^2,$$

where $\tilde{\mathbf{X}}_p$, $p = 0, 1, \cdots, P-1$ are the noisy slabs. It follows that the conditional least squares update for $\mathbf{S}$ is:

$$\hat{\mathbf{S}} = \begin{bmatrix} \hat{\mathbf{A}}\mathbf{D}_0(\hat{\mathbf{H}}) \\ \vdots \\ \hat{\mathbf{A}}\mathbf{D}_{P-1}(\hat{\mathbf{H}}) \end{bmatrix}^{\dagger} \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix},$$

where $(\cdot)^{\dagger}$ stands for pseudo-inverse, and $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$ denote previously obtained estimates of $\mathbf{A}$, and $\mathbf{H}$. Similarly, from the second way of slicing the 3-D data, the conditional LS update for $\mathbf{H}$ is:

$$\widehat{\mathbf{H}}^T = \begin{bmatrix} \widehat{\mathbf{S}}^T\mathbf{D}_0(\widehat{\mathbf{A}}) \\ \vdots \\ \widehat{\mathbf{S}}^T\mathbf{D}_{K-1}(\widehat{\mathbf{A}}) \end{bmatrix}^{\dagger} \begin{bmatrix} \widetilde{\mathbf{Y}}_0 \\ \vdots \\ \widetilde{\mathbf{Y}}_{K-1} \end{bmatrix}.$$

Finally, from the third way of slicing the data, it follows that the conditional LS update for $\mathbf{A}$ is:

$$\widehat{\mathbf{A}}^T = \begin{bmatrix} \widehat{\mathbf{H}}\mathbf{D}_0(\widehat{\mathbf{S}}^T) \\ \vdots \\ \widehat{\mathbf{H}}\mathbf{D}_{N-1}(\widehat{\mathbf{S}}^T) \end{bmatrix}^{\dagger} \begin{bmatrix} \widetilde{\mathbf{Z}}_0 \\ \vdots \\ \widetilde{\mathbf{Z}}_{N-1} \end{bmatrix}.$$

Note that the conditional update of any given factor, as prescribed above, may either improve, or maintain, but cannot worsen the current fit. Globally monotone convergence follows directly from this observation.

2.2. Monte-Carlo Results

In order to test the proposed methodology, we replicated the experimental setup in the simulation results reported in [1]. Even though (in contrast to [1]) we simultaneously estimate (using COMFAC [15]) the signals, azimuth-elevation parameters, and array response, we focus on reporting azimuth-elevation parameter estimates and associated performance results, in part because we can compare with [1], but also in the interest of brevity.

The first experiment is designed to assess performance in a high SNR scenario. Figure 2 presents azimuth-elevation scatter for 4 sources with (azimuth, elevation) = $\{(10,25),\ (15,20)\ ,(20,15),\ (25,10)\}$ degrees, using the proposed method on a 5×5 square array, $\lambda/4$ sensor spacing, three maximal-overlap 16-element subarrays, $N = 100$ symbols, and SNR = 54 dB. Figure 3 presents the same for a 10×10 array, three maximal-overlap 81-element subarrays, $N = 300$ symbols, and SNR = 12 dB. Numerical results for both experiments are summarized in tabular format in Figure 1, which compares the proposed method versus the three methods reported in [1] (M1, M2, and M3), in terms of the mean and standard deviation of the respective estimates. One may observe that the proposed algorithm provides better estimates, both in terms of bias, and in terms of standard deviation. Specifically, the standard deviation performance of the proposed algorithm is two orders of magnitude better than that of the competing algorithms in the high SNR case; and about one order of magnitude better in the low SNR case. The quality of the estimates afforded by the proposed algorithm is also evident from the scatter diagrams in Figures 2, 3.

We have also simulated the experimental setup in example 1 of [2]. This example entails $M = 2$ sources, with (azimuth, elevation) = $\{(-10,10),\ (-5,7)\}$ degrees, a 5×5 square array, maximal-overlap 16-element subarrays, $\lambda/2$ sensor spacing, SNR of 10 dB for each user, 250 trials, and $N = 25,\ 50,\ 100,\ 250$ symbol snapshots. Note that we have used only four maximal-overlap 16-element subarrays, whereas the best method presented in [2] (WSF - weighted subspace fitting, initialized using approximate SSF) implicitly assumes a fully calibrated array, in the sense that: (i) all array elements are assumed to be identical (thus utilizing the finest invariance-granularity possible: twenty-five identical one-element subarrays), and (ii) the optimal weighting was utilized in the WSF search.

Figures 4, 5 present plots of azimuth, elevation RMS error, respectively, for the proposed method (solid with circles), versus the best [WSF initialized by SSF] method reported in [2] (dashed with stars). The two methods provide comparable results: the proposed method gives worse elevation estimates, but slightly better azimuth estimates. The proposed method is better than all other methods reported in [2], including approximate SSF. Note that the proposed method utilizes *only four array invariances*, whereas [WSF initialized by SSF] in [2] uses *twenty-five array invariances*, plus optimal weighting.

Figures 6, 7 present plots of azimuth, elevation RMS error, respectively, versus signal correlation. The setup is as above, but this time N is fixed to 100 snapshots. This scenario corresponds to example 2 in [2]. Again, the proposed method provides results comparable to the best [WSF initialized by SSF] results in [2] (both methods are very robust with respect to signal correlation), but the former only utilizes four array invariances and is otherwise array-blind, whereas the latter utilizes twenty-five invariances, and effectively assumes a calibrated array. We also note that the current algorithm always extracts the azimuth information first, then uses it to extract the elevation parameters. Some further improvement can be obtained by solving the associated system of equations to extract all parameters at once.

3. Discussion and Conclusions

The deterministic approach and link to PARAFAC analysis presented herein offer: (i) A powerful identifiability result; (ii) Support for very small sample sizes (down to $N = 2$): can accommodate very fast variation in fading / azimuth-elevation parameters; (iii) Support for more sources than number of sensors per subarray; (iv) Optimal pairing of azimuth-elevation parameters (no need for approximate/suboptimal matching based on, e.g., matrix perturbations); (v) Simultaneous estimation of the transmitted signals; (vi) LS / deterministic ML interpretation. In addition, the deterministic/ALS approach is robust with respect to statistical signal correlations, and also allows one to incorporate finite-alphabet / constant-modulus constraints into the basic iteration. This can help a great deal in hard-to-resolve cases - the results will be reported elsewhere.

Note that the multiple-parameter / multiple-invariance subspace fitting (SSF) formulation of ESPRIT (e.g., cf. [2, 5, 6, 7]; related work on closed-form (suboptimal) multiple-invariance / multiple parameter ESPRIT appears in [3, 4]) is actually a special case of the mathematical statement of the problem considered in this paper - posed in eigen-space. In particular, given P subspace estimates $\widehat{\mathbf{E}}_0, \cdots, \widehat{\mathbf{E}}_{P-1}$ the multiple-parameter / multiple-invariance ESPRIT SSF problem is to find a square invertible matrix $\mathbf{T}$, unitary diagonal matrices $\mathbf{\Phi}_0, \cdots, \mathbf{\Phi}_{P-1}$, and a (sub)array response matrix $\mathbf{A}$, such that the following cost is minimized:

$$\left\| \begin{bmatrix} \widehat{\mathbf{E}}_0 \\ \vdots \\ \widehat{\mathbf{E}}_{P-1} \end{bmatrix} \mathbf{W}^{\frac{1}{2}} - \begin{bmatrix} \mathbf{A}\mathbf{\Phi}_0 \\ \vdots \\ \mathbf{A}\mathbf{\Phi}_{P-1} \end{bmatrix} \mathbf{T} \right\|_F^2$$

It follows that the identifiability results / trilinear ALS techniques presented herein carry over to the above SSF problem, and its many applications (including multidimensional harmonic retrieval). These are currently under investigation; see also [14].

References

[1] A-J. van der Veen, P.B. Ober, and E.F. Deprettere, "Azimuth and Elevation Computation in High Resolution DOA Estimation", *IEEE Trans. Signal Processing*, 40(7):1828–1832, July 1992.

[2] A. Swindlehurst, and T. Kailath, "Algorithms for Azimuth/Elevation Direction Finding Using Regular Array Geometries", *IEEE Trans. AES*, 29(1):145–156, Jan. 1993.

[3] M. Haardt, M.D. Zoltowski, C.P. Mathews, and J.A. Nossek, "2D Unitary ESPRIT for Efficient 2D Parameter Estimation", In *Proc. ICASSP95*, 3:2096–2099, May 8-12, 1995, Detroit, Michi gan.

[4] K.T. Wong, and M.D. Zoltowski, "Closed-Form Multi-Dimensional Multi-Invariance ESPRIT", In *Proc. ICASSP97*, 5:3489–3492, April 21-24, 1997, Munich, Ger many.

[5] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple Invariance ESPRIT", *IEEE Trans. Signal Processing*, 40(4):868–881, April 1992.

[6] M. Viberg, and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting", *IEEE Trans. Signal Processing*, 39(5):1110–1121, May 1991.

[7] M. Viberg, B. Ottersten, and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting", *IEEE Trans. Signal Processing*, 39(11):2436–2449, Nov. 1991.

[8] J.B. Kruskal, "Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Application to Arithmetic Complexity and Statistics", *Linear Algebra and Its Applications*, 18:95–138, 1977.

[9] J.B. Kruskal, "Rank Decomposition, and Uniqueness for 3-way and N-way Arrays", In *Multiway Data Analysis*, R. Coppi and S. Bolasco (*Eds.*), pp. 7–18, North-Holland, Amsterdam, 1988 (Proc. MULTIWAY'88).

[10] R.A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis", *UCLA Working Papers in Phonetics*, 16:1–84, (1970).

[11] R. Bro, "PARAFAC: Tutorial & Applications", *Chemometrics and Intelligent Laboratory Systems*, 38:149–171, 1997.

[12] L. de Lathauwer, "Signal Processing based on Multilinear Algebra", Doctoral Thesis, Katholieke Universiteit Leuven, Sept. 1997.

[13] P. Comon, "Blind Channel Identification and Extraction of More Sources than Sensors", in *Proc. SPIE San Diego*, July 19-24, 1998.

[14] N.D. Sidiropoulos, G.B. Giannakis, and R. Bro, "Deterministic Waveform-Preserving Blind Separation of DS-CDMA Signals Using an Antenna Array" *Proc. SSAP98*, September 14-16, 1998, Portland, Oregon.

[15] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "A Fast Least Squares Algorithm for Separating Trilinear Mixtures", *ICA'99 - Int. Workshop on Independent Component Analysis and Blind Signal Separation*, Jan. 11-15, 1999, Aussois, France (to appear).

Method	Source number	SNR = 54dB, 5x5 array, N = 100 symbols				SNR = 12dB, 10x10 array, N = 300 symbols			
		Mean Azim.	Mean Elev.	Std Azim.	Std. Elev.	Mean Azim.	Mean Elev.	Std. Azim.	Std. Elev.
Propsd	1	10.0002	25.000	0.0037	0.0046	10.0047	25.0001	0.1208	0.1381
	2	15.0002	20.0003	0.0038	0.0043	15.0072	20.0014	0.1255	0.1286
	3	19.9990	14.9998	0.0039	0.0048	19.9976	15.0027	0.1332	0.1372
	4	24.9999	9.9998	0.0039	0.0041	24.9957	9.9898	0.1329	0.1439
M1	1	10.00	24.98	0.2	0.2	10.01	24.96	0.6	1.0
	2	15.01	20.00	0.5	0.6	14.85	19.95	1.3	1.5
	3	19.98	15.01	0.5	0.5	20.07	15.08	1.3	1.5
	4	25.02	9.98	0.2	0.2	25.02	9.99	0.6	0.6
M2	1	9.96	24.99	0.7	0.7	10.11	24.97	0.7	0.8
	2	14.98	19.97	0.7	0.7	15.08	19.97	0.9	0.9
	3	19.98	14.99	0.7	0.7	20.03	15.01	0.9	0.9
	4	25.05	9.92	0.8	0.7	24.90	10.10	0.8	0.7
M3	1	10.02	25.01	0.3	0.3	10.15	24.84	0.8	0.9
	2	15.03	19.99	0.5	0.5	15.04	19.99	1.2	1.2
	3	19.94	15.08	0.4	0.4	19.98	15.04	1.1	1.1
	4	25.01	10.02	0.3	0.3	24.86	10.14	0.8	0.8
M1, M2, M3: Results reported from van der Veen et al., "Azimuth and Elevation Computation in High-Resolution DOA Estimation", IEEE Trans. Signal Processing, 40(7):1828-1832, July 1992.									

Figure 1. Estimates of Statistics - see text.

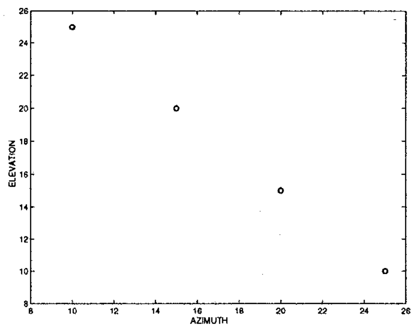

Figure 2. Elevation vs. Azimuth scatter, SNR = 54 dB (see text).

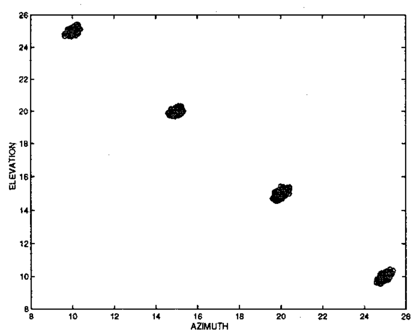

Figure 3. Elevation vs. Azimuth scatter, SNR = 12 dB (see text).

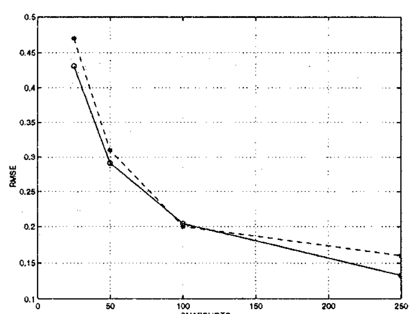

Figure 4. Azimuth RMSE vs. snapshots (N). Solid with ∘ = proposed; dashed with * = SSF + WSF method in [2] (see text).

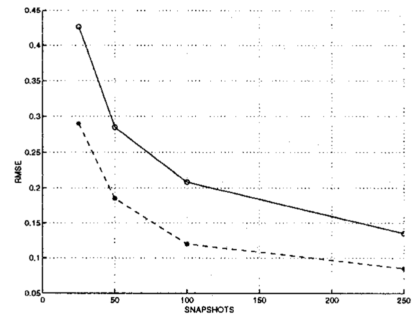

Figure 5. Elevation RMSE vs. snapshots (N). Solid with ∘ = proposed; dashed with * = SSF + WSF method in [2] (see text).

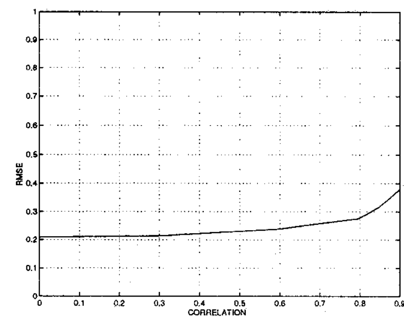

Figure 6. Azimuth RMSE vs. signal correlation.

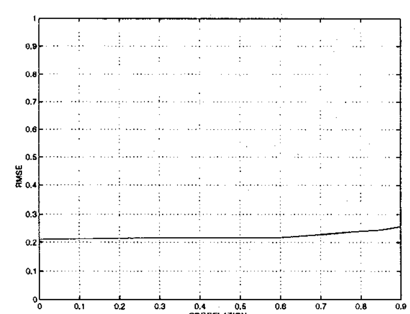

Figure 7. Elevation RMSE vs. signal correlation.