TOWARDS A NEW MATRIX DECOMPOSITION

Jean-Jacques FUCHS

IRISA/Université de Rennes I
Campus de Beaulieu - 35042 Rennes Cedex - France
fuchs@irisa.fr

ABSTRACT

Low rank matrix approximations have many applications in different domains. In system theory it has been used in model reduction schemes, in system identification with output-error models and in static errors-in-variables problems, for instance. The approximations are mostly performed using the singular value decomposition. This is optimal for all unitarily invariant matrix norms, such as the Frobenius norm. From a statistical point of view it is justified when the components are perturbed by independent and identically distributed zero mean Gaussian noise. If this assumption is not valid other norms and thus approximations should be considered. Below we consider the l1-norm that is optimal if the noise samples follow the Laplace or double-exponential distribution and we indicate how to obtain for an arbitrary matrix, the optimal decomposition -similar to the singular value decomposition- associated with this norm.

1. INTRODUCTION

Approximating a given matrix by a matrix of lower rank is of interest in many areas of systems theory, signal or image processing. If the components of a real (m,n)-dimensional matrix A of rank r are corrupted by independent identically distributed zero-mean random variables it will generically become full rank and estimating the true underlying matrix is a challenging problem that may require a preliminary estimation of the rank. In case the noise is Gaussian the norm to use in the approximation procedure is the so-called Frobenius norm, we denote $|||A\,|||_2$:

$$|||A\,|||_2 = (\sum_{i,j} a_{i,j}^2)^{1/2} = \text{trace}(A^T A)^{1/2} \qquad (1)$$

where A^T denotes the transpose of the matrix A. In case the true rank is known, the problem to solve is

$$\min_{\Delta} \; |||\Delta\,|||_2 \quad \text{s.t.} \quad \text{rank}(A-\Delta) = r \qquad (2)$$

Its optimum is well known [1, 2] and easily deduced from the singular value decomposition (SVD) of the matrix A. The SVD has become an important tool in many areas of signal processing precisely because of its link with this and similar matrix approximation properties.

In case the noise that corrupts the components is non-Gaussian but is assumed to follow a double-exponential or Laplace distribution $\frac{\beta}{2}\, e^{-\beta|x|}$ the matrix norm to be used is

$$|||A\,|||_1 = \sum_{i,j} |a_{i,j}|. \qquad (3)$$

Indeed since the noises are assumed to be independent on the different components, the joint density is simply the product of the common density and maximizing the log-likelihood leads to minimizing the Frobenius norm in the Gaussian case and this 'ℓ_1'-norm in the Laplacian case.

Note that if real (m,n)-matrices are considered as vectors belonging to vector spaces of dimension $n \times m$, the vector norms can become matrix norms provided they satisfy the axioms required for a matrix norm. Both the Euclidean or ℓ_2-norm $\|x\|_2 = (\,\sum_i x_i^2\,)^{1/2}$ and the ℓ_1-norm $\|x\|_1 = \sum_i |x_i|$ are such vector norms that lead to matrix norms, respectively (1) and (3).

To our knowledge the approximation problem (2) associated with matrix norm (3) has not been considered and in this contribution we present some partial results towards its solution.

In section 2, we recall briefly some results concerning the SVD decomposition, its link with the Frobenius norm and some related matrix approximation problems. In section 3 we introduce the new 'ℓ_1'-norm decomposition and an exhaustive search algorithm that allows to build it. We present an example in section 4 before we conclude.

2. SINGULAR VALUE DECOMPOSITION

2.1. Singular Value Decomposition

For the sake of completeness and for later reference we remember that the SVD [1, 2, 3] of the real (m,n)-matrix A of rank r is of the form $A = U\Sigma V^T$ with U and V square unitary matrices and

$$\Sigma = U^T A V = \begin{bmatrix} \bar{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \text{ where } \bar{\Sigma} = \text{diag}(\sigma_1,\, \dots, \sigma_r)$$
$$\text{with } \sigma_1 \geq \sigma_2 \dots \geq \sigma_r > 0.$$

One can further decompose U into $[\ U_1\ U_2\ ]$ with U_1 having r columns and V similarly into $[\ V_1\ V_2\ ]$ to rewrite the SVD in the form

$$A = U_1 \bar{\Sigma} V_1^T = \sum_1^r \sigma_k u_k v_k^T \quad (4)$$

which is sometimes called the singular value factorization. The columns of the matrix U_1 form an orthonormal basis of $\Re(A)$, the range or column space of A while the columns of V_1 form an orthonormal basis of $\Re(A^T)$. The columns of the matrices U_2 and V_2 are then arbitrary bases for the orthogonal complements of $\Re(A)$ and $\Re(A^T)$ respectively. This full-rank factorization of A leads to a simple expression for A^+, the pseudo-inverse of A, $A^+ = V_1 \bar{\Sigma}^{-1} U_1^T$ which is just a specific generalized inverse of A. Remember that A^- is a generalized inverse of A if $AA^-A = A$.

2.2. Links with the Frobenius norm

Let us first establish the following well known [1, 2, 4] matrix approximation result.

Theorem 1: Given a real (m,n)-dimensional matrix A of rank r the solution of

$$\min_{\Delta} \ |||\Delta\,|||_2 \ \text{ s.t. } \ \text{rank}(A - \Delta) = \text{rank}(A) - 1 \quad (5)$$

has value σ_r and is attained by $\Delta = \sigma_r u_r v_r^T$, the "smallest" singular triplet of A (4). □

We prove this well known result since a similar approach will be used below when the Frobenius norm is replaced by the 'ℓ_1' norm. We will assume that the optimum Δ is a rank one matrix. While this is true and quite easy to establish for the Frobenius norm, it is not true or may be difficult to establish for other norms. We will use the following lemma and corollary that are established in the appendix.

Lemma 1: Any rank one matrix Δ that is such that rank(A-Δ) = rank(A) -1, is of the form

$$\Delta = \frac{Axy^TA}{y^TAx} \quad \text{with} \quad y^TAx \neq 0. \quad □$$

Corollary 1: The rank one matrix Δ can also be rewritten

$\Delta = \frac{ab^T}{b^TA^-a}$ with $b^TA^-a \neq 0$, $a \in \Re(A)$, $b \in \Re(A^T)$ and A^- any generalized inverse of A. □

Proof of Theorem 1: Using Lemma 1, the optimization problem (5) can be rewritten

$$\min_{x,y} \ |||\frac{Axy^TA}{y^TAx}\,|||_2 \ \text{ with } y^TAx \neq 0.$$

Since $|||ab^T\,|||_2 = \|a\|_2\,\|b\|_2$, the solution of this problem can in turn be deduced from the optimum of

$$\max_{x,y} \ |y^TAx| \quad \text{with} \quad \|Ax\|_2^2 = 1, \ \ \|A^Ty\|_2^2 = 1.$$

We remove the absolute values in the cost function since x and y are not sign constrained and at the optimum $y^TAx > 0$. The Lagrangian of this problem is then

$$\ell(x, y, \lambda, \mu) = -y^TAx + \tfrac{\lambda}{2}(\|Ax\|_2^2 - 1) + \tfrac{\mu}{2}(\|A^TY\|_2^2 - 1).$$

An optimum of the problem is a stationary points of the Lagrangian [5] and satisfies

$$\begin{array}{ll} -A^Ty + \lambda A^TAx = 0, & \|Ax\|_2 = 1 \\ -Ax + \mu AA^Ty = 0, & \|A^Ty\|_2 = 1. \end{array}$$

Premultiplying by the first relation by x^T (respectively the second by y^T), one observes that at a stationary points of interest $\lambda = \mu = y^TAx > 0$. Introducing then $u = Ax$, $v = A^Ty$ and $\sigma = 1/\lambda = 1/\mu > 0$ these conditions become after some manipulations

$$\begin{array}{lll} A^Tu = \sigma v, & \|u\|_2 = 1, & \sigma > 0 \\ Av = \sigma u, & \|v\|_2 = 1 & \end{array} \quad (6)$$

in which one recognizes the coupled equation defining the singular value triplets $\{\sigma_k, u_k, v_k\}$ also known as the Schmidt pairs [3]. Since $y^TAx = 1/\sigma$ has to be maximized one retains the minimal triplets $\{\sigma_r, u_r, v_r\}$. It is straightforward to verify that the second order sufficient conditions for a strict optimum are satisfied by u_r and v_r if the smallest singular value σ_r is simple. □

Note that we have only proven that

$$\min_{\text{rank}(\Delta)=1} \ |||\Delta\,|||_2 \ \text{ s.t. } \ \text{rank}(A - \Delta) = \text{rank}(A) - 1$$

is attained for $\Delta = \sigma_r u_r v_r^T$ provided σ_r the minimal nonzero singular value of A is unique. Of course replacing now A by $A - \sigma_r u_r v_r^T$ in this problem would yields $\Delta = \sigma_{r-1} u_{r-1} v_{r-1}^T$ and so on. The whole SVD can progressively be constructed by solving this sequence of optimization problems.

We now perform the same reasoning using the 'ℓ_1'-norm (3) in place of the Frobenius norm(1).

3. NEW MATRIX DECOMPOSITION

3.1. Elementary step

We are given a real (m,n)-dimensional matrix A of rank r and we seek the solution of

$$\min_{\text{rank}(\Delta)=1} \ |||\Delta\,|||_1 \ \text{ s.t. } \ \text{rank}(A - \Delta) = \text{rank}(A) - 1. \quad (7)$$

Since we have not been able to establish that the optimum is unique and attained at a rank one matrix, except in the case where A is full rank, we impose this condition. Using

Corollary 1 and the fact that $|||ab^T|||_1 = \|a\|_1 \; \|b\|_1$, the problem (7) is equivalent to

$$\max_{a,b} \; b^T A^- a \;\text{ with }\; \|a\|_1 = 1,\; \|b\|_1 = 1, \\ a \in \Re(A),\; b \in \Re(A^T).$$

We now replace the condition $a \in \Re(A)$ by $U_2^T a = 0$ and $b \in \Re(A^T)$ by $V_2^T b = 0$ where the matrices U_2 and V_2 were introduced below (4). Their columns are arbitrary bases for the orthogonal complements of $\Re(A)$ and $\Re(A^T)$ respectively. Problem (7) is thus equivalent to

$$\max_{a,b} \; b^T A^- a \quad \text{s.t.} \quad \|a\|_1 = 1,\; \|b\|_1 = 1, \\ U_2^T a = 0,\; V_2^T b = 0. \tag{8}$$

The Lagrangian of the problem is defined as

$$\ell(a, b, \lambda_1, \lambda_2, \pi_1, \pi_2) = -b^T A^- a + \lambda_1(\|a\|_1 - 1) \\ + \lambda_2(\|b\|_1 - 1) + \pi_1^T U_2^T a + \pi_2^T V_2^T b$$

with λ_1 and $\lambda_2 \in \mathbb{R}$, $\pi_1 \in \mathbb{R}^m$ and $\pi_2 \in \mathbb{R}^n$.

Due to the presence of the ℓ_1 norm, the gradient does not exist at every point and in order to characterize the conditions satisfied by the stationary points of the Lagrangian, we introduce the sub-differential [5] of $\|z\|_1$ denoted $\partial \, \|z\|_1$, a set of vectors called the sub-gradients, an extension of the gradient

$$\partial \, \|z\|_1 = \{v | v^T z = \|z\|_1,\; \|v\|_\infty \leq 1\} \tag{9}$$
$$= \{v | v_i = \text{sign}(z_i) \;\text{ if }\; z_i \neq 0 \;\text{ and }\; |v_i| \leq 1 \;\text{ otherwise}\}$$

where $\text{sign}(z_i) = 1$ if $z_i > 0$ and $\text{sign}(z_i) = -1$ if $z_i < 0$. The stationary points of the Lagrangian satisfy :

$$-(A^-)^T b + \lambda_1 u + U_2 \pi_1 = 0, \\ -A^- a + \lambda_2 v + V_2 \pi_2 = 0, \\ \|a\|_1 = 1, \|b\|_1 = 1, U_2^T a = 0,\; V_2^T b = 0$$

for some $u \in \partial\|a\|_1$ and $v \in \partial\|b\|_1$.
Since $a^T u = \|a\|_1$ for any $u \in \partial\|a\|_1$ and $a^T U_2 = 0$, premultiplying the first relation by a^T, gives $\lambda_1 = a^T (A^-)^T b$, premultiplying the second relation by b^T one obtains similarly $\lambda_2 = b^T A^- a$. One thus has $\lambda_1 = \lambda_2 = b^T A^- a$, a quantity that is strictly positive for the stationary points of interest since we seek a maximum of $b^T A^- a$.
Introducing $\sigma = 1/\lambda_1 = 1/\lambda_2 > 0$ and observing that $AA^- a = a$ since $a \in \Re(A)$ and the corresponding property for b premultiplying the two relations by A^T and A respectively the necessary conditions for an optimum become

$$A^T u = \sigma b,\;\; \|a\|_1 = 1,\; u \in \partial\|a\|_1,\; \sigma \geq 0 \\ A\, v = \sigma a, \quad \|b\|_1 = 1,\; v \in \partial\|b\|_1. \tag{10}$$

The vectors $a,\; b$ satisfying these conditions associated with the smallest σ and the associated rank one matrix $\sigma a b^T$ is then the optimum of (7). We have not been able so far to find the sufficient conditions this point needs to satisfy.

The relations (10) have to be compared to (6) that characterize the singular triplets of the matrix A. While relations (6) are easily transformed into those defining the eigenvectors and eigenvalues of the matrices AA^T and $A^T A$, here no such combination seems feasible.

Because the vectors u and v are only known to belong to sets these relations are difficult to exploit and to solve them we need the following lemma.

Lemma 2: The optimal rank one matrix $\Delta = \frac{ab^T}{b^T A^- a} = \sigma a b^T$ deduced from the solution to problem (8) is, the outer product of a column vector and a row vector having each at least r-1 zero components. □

Proof : We establish equivalently that an optimal solution say (a_o, b_o) to (8) is such that a_o has at most m-(r-1) non-zero components and b_o at most n-(r-1) non-zero components. This follows by observing that for the optimal b_o vector, the vector a is the optimum of the following optimization problem

$$\max_a b_o^T A^- a \quad \text{subject to} \quad \|a\|_1 = 1,\; U_2^T a = 0$$

This problem can be reformulated as a linear program in standard form, its optimum is then generically attained at a basic feasible solution. Retransposed into the current formulation this says that a_o, the optimal a, has generically at most m-$r + 1$ non-zero components. A similar reasoning leads to the corresponding result for b. □

3.2. Exhaustive search algorithm

Using the results of Lemma 2, we develop an exhaustive algorithm that searches all the solutions (σ, a, b, u, v) that satisfy (10) for A a (m,n) matrix of rank r. The optimum of (7) is then $\Delta = \sigma a b^T$ with σ minimal.

To get all the solutions of (10), we consider all the square order-(r-1) matrices that can be extracted from A. To ease the exposition we assume that for each such square order-(r-1) matrix we exchange the rows and columns in A to bring it to the top left corner and denote it $A_{1,1}$. Below we describe the different steps that then lead to an admissible solution of (10) As soon as the results of one of these steps is not compatible with (10) one *stops* and proceeds to the next square order-(r-1) matrix.

With $A_{1,1}$ is associated a partition of A into 4 blocks. We partition accordingly vector a into a_1 and a_2 with a_1 its first r-1 components and do the same for b, u and v.

We assume that for the current partition $a_1 = b_1 = 0$, see Lemma 2 . The remaining components in a_2 and b_2 are then generically non-zero which in turn implies (9) that the components in u_2 and v_2 are such that u_2=sign(a_2) and v_2=sign(b_2).

Using these notations, the second relation in (10) leads to two block equations

$$\begin{aligned} A_{1,1}v_1 + A_{1,2}\,\text{sign}(b_2) &= \sigma a_1 = 0 \\ A_{2,1}v_1 + A_{2,2}\,\text{sign}(b_2) &= \sigma a_2. \end{aligned}$$

If $A_{1,1}$ is not invertible one *stops*, otherwise the first of these equations allows to compute the vector v_1 associated with each of the $2^{(n-r)}$ potential vectors $\text{sign}(b_2) \in [\pm 1 \;\; \pm 1 \;\; .. \pm 1\;]^T$ that need to be considered. For each of these $\text{sign}(b_2)$-vectors one checks if the associated v_1-vector satisfies $\|v_1\|_\infty \leq 1$ and if this does not hold one proceeds to the next $\text{sign}(b_2)$-vector. Otherwise one has a coherent couple $\{v_1,\ \text{sign}(b_2)\}$ which when substituted in the second equation equation gives $\sigma > 0$ and a_2 with $\|a_2\|_1 = 1$. One then proceeds with the current subset of unknowns $\{a_2, v_1,\ \text{sign}(b_2), \sigma\}$ to the two equations drawn from $u^T A = \sigma b^T$ the transpose of the first relation in (10)

$$\begin{aligned} u_1^T A_{1,1} +\ \text{sign}(a_2)^T A_{2,1} &= \sigma b_1^T = 0 \\ u_1^T A_{1,2}\ +\ \text{sign}(a_2)^T A_{2,2} &= \sigma b_2^T. \end{aligned}$$

Since $\text{sign}(a_2)$ is known, one draws u_1 from the first equation above and checks if $\|u_1\|_\infty \leq 1$. If this condition is not satisfied one *stops*, if it holds one deduces from the second equation $\sigma' > 0$ and b_2 with $\|b_2\|_1 = 1$. If $\sigma' = \sigma$ and if the signs of the components in b_2 are identical to $\text{sign}(b_2)$ already obtained we have found a stationary point of the Lagrangian, i.e., an admissible solution (σ, a, b, u, v) of (10).

There are in general quite few admissible sets and their number is not fixed and predictable as is the case when one solves (6). We retain the set associated with the smallest σ that is generically unique.

There is another way to implement an exhaustive algorithm that leads to the same solution. The idea is again to consider all the square order-(r-1) matrices that can be extracted from A and to achieve the permutations that bring these sub-matrices to the top left corner where we denote it $A_{1,1}$.

The above development can be interpreted as follows. For each $A_{1,1}$ one seeks the perturbation $\Delta = \sigma ab^T$ which is zero except for its block $\Delta_{2,2}$ that makes the rank of A drop by one, when subtracted from A. For each $A_{1,1}$, the elements of the perturbation $\Delta_{2,2} = \sigma a_2 b_2^T$ are computed step by step and one *stops* as soon as a step is not compatible with the constraints imposed by (10).

The other way to get the same solution is then as follows. The perturbation Δ that is zero except for $\Delta_{2,2}$ exists if and only if $A_{1,1}$ is invertible. Its value is $\Delta_{2,2} = A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2}$ which can be seen as the Schur complement of $A_{1,1}$ in A. One can prove that it is a rank one matrix which can thus always be written $\sigma a_2 b_2^T$ with $\|a_2\|_1 = 1$, $\|b_2\|_1 = 1$ and $\sigma > 0$. But the so-obtained triplet σ, a_2, b_2 is seldom coherent with the constraints imposed by (10). The rare cases where this is true are the same as those given by the previous algorithm.

4. EXAMPLE

For illustration purposes we present the decomposition of a matrix A that is obtained by repeatedly applying the above algorithm first to A then to $A - \sigma ab^T$ and so on.

For a square invertible matrix of order n, Lemma 2 tells us that the least 'l_1' norm matrix $\Delta_1 = \sigma_1 a_1 b_1^T$ that makes its rank drop to $n-1$ is a matrix having just one non-zero element. This solution to this specific optimization is easy to obtain directly. The non-zero element is located symmetrically to the place where A^{-1} attains its maximum (see corollary 1). Having subtracted the corresponding minimal 'l_1' norm matrix from A, the resulting matrix is of rank $n-1$ and lemma 2 tells us that the least 'l_1' norm matrix of rank one $\Delta_2 = \sigma_2 a_2 b_2^T$ that makes its rank drop to $n-2$ is obtained has the outer product of two vectors, say $\sigma_2 a_2$ and b_2, having each at most two non-zero elements. Proceeding this way the k-th rank one matrix in the decomposition is the product of two vectors having each at most k non-zero elements and so on.

For the Vandermonde matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ .2 & .4 & .6 & .8 \\ .04 & .16 & .36 & .64 \\ .008 & .064 & .216 & .512 \end{bmatrix}$$

the decomposition one obtains is of the form $A = GDH^T = \sum_i \sigma_i a_i b_i^T$ with in the columns of G the unit ℓ_1-norm vectors a_i, in the columns of H the unit ℓ_1-norm vectors b_i and the σ_i's on the diagonal of D

$$G = \begin{bmatrix} 0 & 0 & -0.5938 & -0.4615 \\ 0 & 0.5682 & 0 & -0.2592 \\ -1 & 0 & 0.1781 & -0.1661 \\ 0 & -0.4318 & 0.2280 & -0.1131 \end{bmatrix}$$

$$H = \begin{bmatrix} 0 & -0.7000 & -0.4051 & -0.1411 \\ 1 & 0 & -0.2405 & -0.1950 \\ 0 & 0.3000 & 0 & -0.2739 \\ 0 & 0 & 0.3544 & -0.3900 \end{bmatrix}$$

and D is equal to $\text{diag}(0.0100,\ 0.2247,\ 2.0157,\ 7.9121)$.

5. CONCLUDING REMARKS

We have presented some very preliminary results leading to a new matrix decomposition built around the matrix norm (3) deduced from the ℓ_1 vector norm when real (m,n)-matrices are considered as $(n \times m)$-vectors. It is the decomposition that should be used in place of the SVD when the components of the matrix are observed in the presence of additive Laplacian noise rather than Gaussian noise.

The ℓ_1 norm is often used when either sparseness or robustness is of interest. Sparseness is a pretty active domain of research these days [6, 7]. In this contribution the link between sparseness and the ℓ_1 norm can be seen in the fact that the matrix perturbation Δ we obtain at each step has generically the smallest number of nonzero components, see Lemma 2.

In conjunction with robustness the ℓ_1 norm is used since it is less sensitive to outliers [8] than the usual Euclidean norm and it is certainly more on this side that applications of extensions of the present contribution can be expected. In the presence of outliers the identification of output-error models or model reduction schemes would probably benefit if the usual ℓ_2 norm were replaced by the ℓ_1 norm or better of mixture of both, like Huber's induced cost, which is no longer a norm.

This same motivation leads to the what could be called the Total Least Absolute Deviation problem [9], the equivalent to the well known Total Least Squares (TLS) problem when the Frobenius norm (1) is replaced by (3).

6. REFERENCES

[1] S. Van Huffel and J. Vanderwalle, *The Total Least Squares problem*. SIAM, 1991 .

[2] G.H. Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins University Press, 1983.

[3] P. Dewilde and E.F. Deprettere. "Singular Value Decomposition : an introduction" *SVD and Signal Processing* pp. 3-41, North-Holland, 1988.

[4] G.W. Stewart and J.G. Sun. *Matrix Perturbation Theory.* Academic Press, 1990.

[5] R. Fletcher, *Practical Methods of Optimization*. John Wiley and sons., 1987.

[6] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. on I.T.*, 47, 11, 2845-2862, Nov. 2001.

[7] J.J. Fuchs, "More on sparse representations in arbitrary bases," *IEEE Trans. on I.T.*, 50, 6, 1341-1344, June 2004.

[8] P.J. Huber. *Robust Statistics*. John Wiley and sons., 1981.

[9] M.R. Osborne and G.A. Watson. An analysis of the total approximation problem in separable norms and an algorithm for the total ℓ_1 problem. *SIAM J. Sci. Stat. Comput.*, vol. 6, 2, 410–424, April 1985.

[10] R.B. Bapat, *Linear Algebra and Linear Models*. Springer, Universitext, 1999, p. 114.

7. APPENDIX

Lemma 1 Any rank one matrix Δ that is such that rank(A-Δ) = rank(A) -1, is of the form

$$\Delta = \frac{Axy^TA}{y^TAx} \quad \text{with} \quad y^TAx \neq 0. \qquad \square$$

Proof : To establish that Δ is of the given form, we use the following *Rank additivity characterization* [10]:

rank(A) = rank($A - \Delta$) + rank(Δ)

if and only if any generalized inverse A^- of A is also a generalized inverse of $A - \Delta$.

Remember that A^- is a generalized inverse of A if $AA^-A = A$. Take $\Delta = \alpha ab^T$ with α a scalar and a, b vectors of unit Euclidean norm. The relation $(A-\Delta)A^-(A-\Delta) = A-\Delta$ becomes after substitution of Δ and simplification by α:

$$AA^-ab^T + ab^TA^-A - ab^T - \alpha(b^TA^-a)ab^T = 0 \quad (\dagger)$$

Post-multiply both sides of (†) by b gives $AA^-a = \beta a$ with β a scalar. But this relation tells us that $a = Ax$ for some x which implies that $AA^-a = a$, i.e. $\beta = 1$. Pre-multiply both sides of (†) by a^T to get $b^TA^-A = \gamma b^T$ with γ a scalar. A similar analysis implies $\gamma = 1$. Substituting in (†) one gets $\alpha = \frac{1}{b^TA^-a}$ which can be rewritten $\frac{1}{y^TAx}$ and the announced form follows. $\square$

The Corollary below follows immediately.

Corollary 1: The rank one matrix Δ can also be rewritten

$$\Delta = \frac{ab^T}{b^TA^-a} \quad \text{with} \quad b^TA^-a \neq 0,\ a \in \Re(A),$$

$b \in \Re(A^T)$ and A^- any generalized inverse of A. $\square$