On Optimal Dimension Reduction in Least-Square System Identification

Majid Fozunbal

HP Laboratories
HPL-2008-219

Keyword(s):
Dimension reduction, echo control, least-square methods, manifold learning, and regularization

Abstract:
The least-square optimization problem in multi-channel echo control is severely ill-conditioned. Methods to mitigate this problem by decorrelating input signals result in undesired audio distortion. Recently, we demonstrated this approach can be tackled by dimension reduction [1]. In this paper we extend our results by studying the trade-off between the approximation error, i.e. the error of reducing the dimension of the search space, and estimation error, i.e. the error caused by observation noise, as function of the reduction in dimension. Simple expressions are derived to determine the optimal dimension as a function of signal-tonoise ratio and condition number of the normal equations.

External Posting Date: December 18, 2008 [Fulltext] Approved for External Publication
Internal Posting Date: December 18, 2008 [Fulltext]

ON OPTIMAL DIMENSION REDUCTION IN LEAST-SQUARE SYSTEM IDENTIFICATION

Majid Fozunbal

Hewlett-Packard Laboratories
Palo Alto, CA 94304, USA
majid.fozunbal@hp.com

ABSTRACT

The least-square optimization problem in multi-channel echo control is severely ill-conditioned. Methods to mitigate this problem by decorrelating input signals result in undesired audio distortion. Recently, we demonstrated this approach can be tackled by dimension reduction [1]. In this paper we extend our results by studying the trade-off between the approximation error, i.e. the error of reducing the dimension of the search space, and estimation error, i.e. the error caused by observation noise, as function of the reduction in dimension. Simple expressions are derived to determine the optimal dimension as a function of signal-to-noise ratio and condition number of the normal equations.

Index Terms— Dimension reduction, echo control, least-square methods, manifold learning, and regularization.

I. INTRODUCTION

Manifold learning [2] is a collection of techniques to estimate a low dimensional (non-linear) model from a set of noisy observations. This low dimensional model is represented as a low-dimensional (linear or non-linear) *manifold* embedded in a higher dimensional linear embedding space [3]. Figure 1 depicts an example where a one-dimensional helix in $\mathbb{R}^3$ models a noisy data set of three dimensional vectors.

In this paper we study manifold learning in the context of *regularizing* least square problems. In particular, we study manifold learning as a method to mitigate the *non-uniqueness problem* in multi-channel echo control (MEC) [1], [4]. MEC systems seek to minimize the energy of echo, typically expressed as a minimization problem of the form

$$\inf_{x\in\mathbb{R}^D} \|Ax-b\|^2. \tag{1}$$

Here, A denotes an excitation matrix formed by the signals sent to the loudspeakers, b is a vector of microphone signals, and x is a candidate estimate of the echo path impulse responses. Because of the spatial correlation of excitation signals, (1) is typically under-determined and ill-conditioned.

A variety of approaches have been proposed to tackle this problem, mostly by making A better conditioned using non-linear and time-varying decorrelation techniques [5], [6]. The major drawback of these methods is the incurred distortion of the audio signals: effectively, echo control quality performance is improved at the cost of other non-linear distortions.

In [1], we proposed *manifold learning* approach to regularize (1), allowing a physically relevant well-conditioned solution. The procedure of this approach is as follows. Let the underlying search space (for impulse responses in a given room) be a d-dimensional linear manifold $\mathcal{M}$ embedded in $\mathbb{R}^D$, where $d < D$. This assumption is supported by experimental observations that show a major portion of the search space lies within a minor portion of dimensions. Since the knowledge of $\mathcal{M}$ is not available, we use a data set of (noisy) samples $\mathcal{D}_n = \{x_1, \cdots, x_n\} \subset \mathbb{R}^D$ to form an approximation $\hat{\mathcal{M}}$ of $\mathcal{M}$. More specifically, given a (noisy) sample $\mathcal{D}_n = \{x_1, \cdots, x_n\} \subset \mathbb{R}^D$, we first compute the empirical correlation $\Lambda = \frac{1}{n}\sum_i x_i x_i'$. Then, an appropriate number d of largest eigenvectors of Λ are computed and used as the basis for $\hat{\mathcal{M}}$. Finally, instead of (1), we solve

$$\inf_{x\in\hat{\mathcal{M}}} \|Ax-b\|^2. \tag{2}$$

The details of this algorithm are given in [1]. In this paper we extend the analysis of [1] by studying the optimal reduction in dimension. We start with an average total error analysis for the solution of (2). We then derive an upper-bound on the average total error. Consequently, we factorize this bound into two additive terms: the *approximation error* and the *estimation error*, respectively. The approximation error is a measure of deviation between $\hat{\mathcal{M}}$ and $\mathcal{M}$, whereas the estimation error measures the deviation due to additive noise and the ill-conditioning of the system matrix A.

We continue by deriving expressions for the trade-off between the approximation and estimation error terms. We find that as the manifold dimension d increases, the estimation error increases linearly, but the approximation error drops more super-linearly. We derive a closed form expression to find the optimal value for d for a minimum total error. It is shown that at this value the marginal decrease in approximation error equals the marginal increase in estimation error. This condition is further expressed in terms of cumulative sum of eigenvalues of the correlation matrix Λ, the signal-to-noise

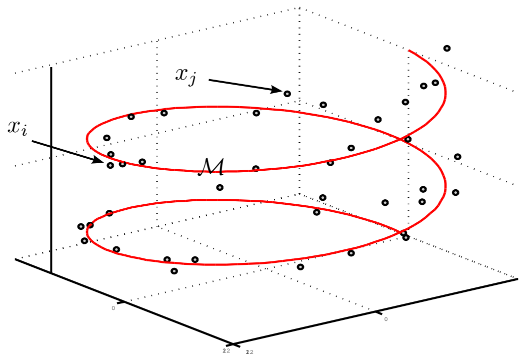

Fig. 1. Example of a one-dimensional helix $(\sin t, \cos t, t)$ that models a noisy collection of points in $\mathbb{R}^3$.

ratio, denoted by γ, and the (aggregated) condition number of $A'A$, denoted by $\chi(A'A)$. In one particular case we find that optimal value for d equals the number of eigenvalues of Λ that are larger than

$$\frac{\chi(A'A)}{\gamma D^2}.$$

In the following sections, we first introduce the technique of linear dimension reduction. We then derive expressions for the approximation error and the estimation error, respectively. Finally, we derive and numerically illustrate the optimality conditions for d.

II. LINEAR DIMENSION REDUCTION

Several techniques exist for manifold learning and dimension reduction [2]. These techniques are divided into two main classes, viz. linear and non-linear techniques. More precisely, given $\mathcal{D}_n$, the problem can be expressed as finding a contiuous[1] *immersion* function

$$f: \mathcal{N} \rightarrow \mathcal{M}$$

whose inverse, referred to as *submersion*,

$$f^{-1}: \mathcal{M} \rightarrow \mathcal{N}$$

is also continuous. Here, $\mathcal{N} \subset \mathbb{R}^d$ and $d < D$ denotes the dimension of manifold. Note that f^{-1} is a projection on the manifold and f is an embedding in the ambient space. Thus, f^{-1} is only a left inverse.

Let define

$$\rho = \frac{d}{D} \tag{3}$$

as the *relative dimension*. Given a relative dimension and a class of candidate functions, the function f is found by minimizing the average error between $x_i \in \mathcal{D}_n$ and its corresponding output $\tilde{x}_i$ resulting from the process, as follows. Let f and f^{-1} be given by

$$f^{-1}(x) \quad = U'x \tag{4}$$

$$f(y) \quad = Uy \tag{5}$$

where U is a $d \times D$ orthogonal matrix such that $U'U = I_d$. The optimal choice for U is found by minimizing

$$\frac{1}{n} \sum_{x_i \in \mathcal{D}_n} \|x_i - UU'x_i\|^2. \tag{6}$$

Solving this equation for U, U is a matrix formed by the d largest eigenvectors of the empirical correlation matrix

$$\Lambda = \frac{1}{n} \sum_{x_i \in \mathcal{D}} x_i x_i'. \tag{7}$$

Let

$$\Lambda = [U \; V] \begin{bmatrix} \Sigma & 0 \\ 0 & \Delta \end{bmatrix} \begin{bmatrix} U' \\ V' \end{bmatrix}.$$

denote the singular value decomposition of Λ where Σ and Δ are diagonal matrices containing the d largest eigenvalues and $D - d$ lowest eigenvalues in descending order. The minimum of (6) is

$$\frac{1}{n} \sum_{x_i \in \mathcal{D}_n} \|V'x_i\|^2 = \operatorname{tr}(\Delta). \tag{8}$$

A key observation obtained from experimental results is that $\operatorname{tr}(\Delta)$ is a small percentage of $\operatorname{tr}(\Lambda)$. Figure 2 depicts the accumulative sum of the eigenvalues of 2000 acoustic impulse responses between random pairs of locations in a $5\times10\times3$ meters room. These impulse response are computed using the image method [7] for a sampling frequency of 8KHz. The x-axis is ρ (3), the relative dimension. The y-axis represents the normalized accumulative sum of the $\lfloor \rho D \rfloor$ eigenvalues, more precisely given as

$$Q(\rho) = \frac{\sum_{i=1}^{\lfloor \rho D \rfloor} \lambda_i}{\sum_{i=1}^{D} \lambda_i} \tag{9}$$

that is a measure of accuracy in approximating Λ by truncating its lowest $D - d$ eigenvalues. In (9), λ_i denotes the i-th largest eigenvalue of Λ.

From Figure 2, it is quite noticeable that about 84% of the norm of Λ lies in 10% of its eigenvalues. Moreover, we conclude from (9) and from Figure 2 that as ρ increases, accuracy improves. However, this increase in approximation accuracy comes at the expense of an increased estimation error, to be discussed below. In Section IV, we derive an expression for the optimal value of ρ.

III. ERROR ANALYSIS

Let denote the approximation for $\mathcal{M}$ by

$$\hat{\mathcal{M}} = \{x \in \mathbb{R}^D : V'x = 0\}. \tag{10}$$

[1]Not necessarily the Euclidean topology.

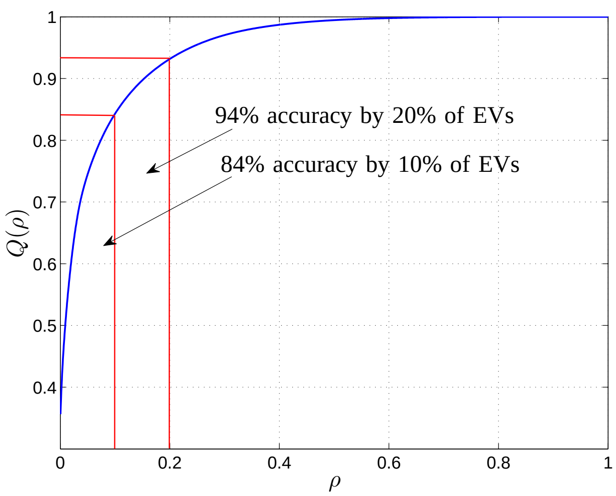

Fig. 2. Cumulative portion of k largest eigenvalues. A number of 2000 impulse responses are generated via a simulation model [Allen & Berkley, '79]. Absorption value is .2, sampling frequency 8 KHz, and room's dimensions are $5 \times 10 \times 3$ meters.

Now, we can express (2) as

$$\begin{aligned} &\inf \|Ax - b\|^2 \\ &\text{s.t. } V'x = 0. \end{aligned} \tag{11}$$

Suppose the true underlying element is x_o, and let

$$b = Ax_o + \nu$$

where ν denotes the additive noise term. Solving (11) for x we need to satisfy

$$(U'A'AU)U'x = U'A'Ax_o + U'A'\nu. \tag{12}$$

Figure 3 attempts to depict the geometry of this solution for a hypothetical example in $\mathbb{R}^2$ where $\|\nu\| = 0$. The deviation between the estimated solution x and the true solution x_o can be expanded by

$$\|x - x_o\|^2 = \|V'x_o\|^2 \left(1 + \frac{\|U'(x - x_o)\|^2}{\|V'x_o\|^2}\right).$$

In this equation, the second multiplicative term is a measure of an angular alignment between the estimated manifold $\hat{\mathcal{M}}$ and the space of solutions $A(x - x_o) = 0$. For the example in Figure 3, this term is $1 + \tan^2\theta$. Thus, fixing $\|V'x_o\|$ the minimum error is when $\theta = 0$.

With some algebraic manipulation, it turns out that

$$\begin{aligned} \|x - x_o\|^2 \leq\ & 2\|V'x_o\|^2(1 + \|(U'A'AU)^{-1}U'A'AV\|^2) \\ & + 2\|(U'A'AU)^{-1}U'A'\nu\|^2. \end{aligned}$$

The first term of this inequality is the effect of approximating $\mathcal{M}$ and the second term is the effect of the additive noise.

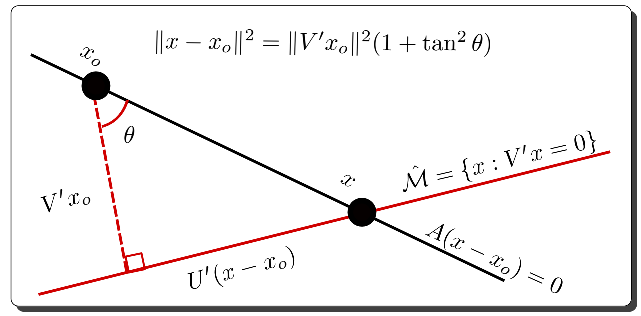

Fig. 3. A hypothetical example to illustrate the geometry of error in terms of the angular alignment between $\hat{\mathcal{M}}$ and the space of solutions $A(x - x_o)$.

Assume that x_o is a random point in $\mathcal{D}_n$ drawn using the empirical density of $\mathcal{D}_n$. Moreover, assume a uniform distribution over all matrices similar[2] to $A'A$. Now, taking the expectation with respect to x_o and $A'A$, we obtain

$$\mathbb{E}\left(\|x - x_o\|^2\right) \lessapprox C(\rho)$$

where

$$\begin{aligned} C(\rho) \triangleq\ & 2(1 - Q(\rho))\mathrm{tr}\,(\Lambda)\left(1 + \mathrm{tr}\left((A'A)^{-1}\right)\rho\right) \\ & + 2\frac{\chi(A'A)}{D\,\gamma}\rho \end{aligned} \tag{13}$$

denotes an upper-bound on total error in terms of *average signal-to-noise ratio* per dimension

$$\gamma = \frac{\mathrm{tr}\,(A'A)}{D\|v\|^2}$$

and the *aggregated condition number* of matrix $A'A$

$$\chi(A'A) = \mathrm{tr}\,(A'A)\,\mathrm{tr}\left((A'A)^{-1}\right).$$

Note that $\chi(A'A)$ is a lower bound of the commonly used notion of condition number that is the ratio of maximum eigenvalue to the minimum eigenvalue.

IV. OPTIMAL DIMENSION

The expression for total (bound on) error $C(\rho)$ has two additive terms. The first term represents (an upper bound) on the average approximation error that occurs by approximating $\mathcal{M}$ with $\hat{\mathcal{M}}$. Let denote this term by

$$C_{\mathrm{app}}(\rho) \triangleq 2(1 - Q(\rho))\mathrm{tr}\,(\Lambda)\left(1 + \mathrm{tr}\left((A'A)^{-1}\right)\rho\right). \tag{14}$$

The other additive term in (13) represents the estimation error caused by the additive noise. The second term is an upper bound on estimation error, denoted by

$$C_{\mathrm{est}}(\rho) \triangleq 2\frac{\chi(A'A)}{\gamma D}\rho. \tag{15}$$

[2]A matrix B is similar to $A'A$ if there exists a non-singular matrix S such that $B = S^{-1}A'AS$.

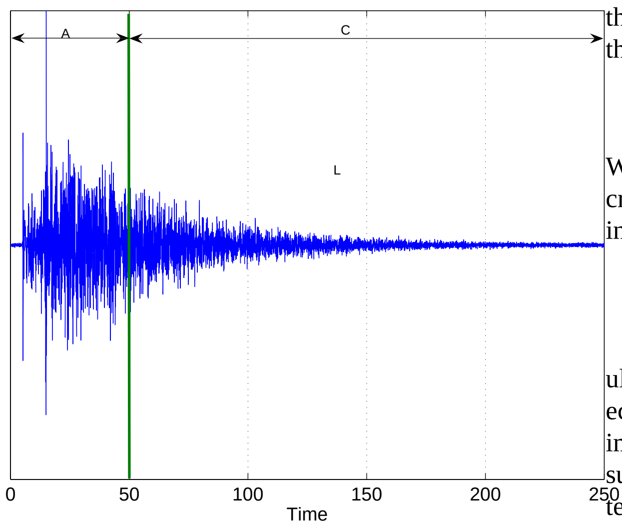

Fig. 4. Approximation error and estimation error and the overall error are depicted for $\gamma = 10$dB, $D = 2000$, and three different values for $\chi(A'A)$. As $\chi(A'A)$ decreases or γ increases the optimal choice for ρ shifts to the right.

Figure 4 demonstrates the behavior of these two error terms for $\gamma = 10$dB and $D = 2000$. As ρ increases, the approximation error $C_{\text{app}}(\rho)$ decreases but the estimation error increases. It is seen that the optimal ρ shifts to the right as γ increases or the condition number of $A'A$ decreases.

The extremum point of (13) can be found by considering ρ as a continuous variable and taking derivatives with respect to ρ. We then obtain the following optimality condition

$$\frac{\partial C_{\text{app}}(\rho)}{\partial \rho} = -\frac{\partial C_{\text{est}}(\rho)}{\partial \rho}. \quad (16)$$

The condition (16) implies that the optimal ρ is the point in which the marginal reduction in approximation error is offset by the marginal increase in estimation error. With some straightforward manipulation of (16), we obtain

$$\frac{\partial Q(\rho)}{\partial \rho} = \frac{1 - Q(\rho) + \frac{\chi(A'A)}{\gamma D \text{tr}(\Lambda)\text{tr}((A'A)^{-1})}}{\rho + 1/\text{tr}\left((A'A)^{-1}\right)}. \quad (17)$$

This equation expresses the condition for optimal choice of ρ with respect to the normalized accumulative sum of the eigenvalues of Λ along with γ and the condition number of $A'A$. It can be solved numerically to find a threshold η such that the number of eigenvalues no smaller than η determines d.

We can further simplify (17) if $\text{tr}\left((A'A)^{-1}\right) \ll 1$. If this condition holds true, (17) boils down to

$$\frac{\partial Q(\rho)}{\partial \rho} \approx \frac{\chi(A'A)}{\gamma D \,\text{tr}\,(\Lambda)} \quad (18)$$

Equivalently, by substituting for $Q(\rho)$ from (9), we obtain the optimal dimension d as the number of eigenvalues of Λ that are no smaller than the cutoff threshold

$$\eta \triangleq \frac{\chi(A'A)}{\gamma D^2}. \quad (19)$$

We conclude that the optimal d increases as $\chi(A'A)$ decreases or as γ increases, a behavior which is also observed in Figure 4.

V. CONCLUSION

We considered optimal linear dimension reduction to regularize ill-conditioned least square problems in multichannel echo control. We observed that a significant part of the impulse-response search space lies in a lower dimensional subspace. This motivated the use of dimension reduction techniques to regularize the echo control problem. We have analyzed the total error in terms of approximation and estimation errors, resulting in a simple expressions for the optimal subspace dimension. In future work we will experimentally verify optimal dimension reduction in actual real-time multi-channel echo control systems.

ACKNOWLEDGEMENT

The author is thankful to Ton Kalker for comments that improved the presentation of the document.

REFERENCES

[1] M. Fozunbal, T. kalker, and R. Schafer, “Multi-channel echo control by model learning,” *Proc. Intern. Workshop Acoustic Echo and Noise Control*, Sept. 2008.

[2] A. Gorban, B. Kegl, D. Wunsch, and A. Zinovyev, *Principal Manifolds for Data Visualisation and Dimension Reduction*. Berlin: Springer, 2007.

[3] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 2002.

[4] M. Sondhi, D. Morgan, and J. L. Hall, “Stereophonic acoustic echo cancellation-an overview of the fundamental problem,” *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148–151, 1995.

[5] J. Bensty, T. Gansler, D. Morgan, M. Sondhi, and S. Gay, *Advances in Network and Acoustic Echo Cancellation*. Springer-Verlag, 2001.

[6] M. Ali, “Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation,” *Proc. IEEE Intern. Conf. Acoustics, Speech, and Signal Proc.*, pp. 3689–3692, 1998.

[7] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 66, no. 4, pp. 943–950, Apr. 1979.