I T H E A

International Journal

INFORMATION THEORIES & APPLICATIONS

2009 Volume 16 Number 2

International Journal
INFORMATION THEORIES & APPLICATIONS
Volume 16 / 2009, Number 2

IJ ITA is official publisher of the scientific papers of the members of the ITHEA® International Scientific Society

IJ ITA welcomes scientific papers connected with any information theory or its application.

IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA as well as the **subscription fees** are given on *www.ithea.org* .
The camera-ready copy of the paper should be received by http://ij.ithea.org.
Responsibility for papers published in IJ ITA belongs to authors.

General Sponsor of IJ ITA is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal "INFORMATION THEORIES & APPLICATIONS" Vol.16, Number 2, 2009

Printed in Bulgaria

Edited by the **Institute of Information Theories and Applications FOI ITHEA®**, Bulgaria,
in collaboration with the V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,
and the Institute of Mathematics and Informatics, BAS, Bulgaria.

Publisher: **ITHEA®**
Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com

ISSN 1310-0513 (printed) **ISSN 1313-0463 (online)** **ISSN 1313-0498 (CD/DVD)**

USING RANDOMIZED ALGORITHMS FOR SOLVING DISCRETE ILL-POSED PROBLEMS

Elena G. Revunova, Dmitri A. Rachkovskij

Abstract: In this paper, we develop an approach for improving the accuracy and speed of the solution of discrete ill-posed problems using the pseudo-inverse method. It is based on a random projection of the initial coefficient matrix. First, a brief introduction is given to least squares and discrete ill-posed problems, approximate matrix decompositions with randomized algorithms, and randomized least squares approximations. Then, we describe the techniques used in this paper to solve discrete ill-posed inverse problems, including the standard Tikhonov regularization and pseudo-inverse after a random projection. The bias-variance decomposition of solution errors is provided for different solution techniques and it is shown experimentally that the error has a minimum at some value of the varied smaller dimension of the random projection matrix. Taking two well-known test examples of the discrete ill-posed problems of Carasso and Baart, we obtain experimental results of their solution. The comparison shows that the minimal error of the pseudo-inverse solution after a random projection is close to the error of the standard Tikhonov regularization, but the former solution is obtained faster, especially at the larger noise values.

Keywords: discrete ill-posed problems, pseudo-inverse, regularization, random projection, bias, variance

ACM Classification Keywords: I.5.4 Signal processing, I.6 SIMULATION AND MODELING (G.3), G.1.9 Integral Equations

Introduction

Many practical applications require solving a discrete problem in the form:

$$\mathbf{Ax} \approx \mathbf{b}. \quad (1)$$

Here the matrix **A** and the vector **b** are known, the vector **x** must be evaluated. In case when

- the singular values of **A** decay gradually to zero,

- the ratio between the largest and the smallest nonzero singular values is large,

the problem is known as discrete ill-posed problem [Hansen, 1998].

Approximate solutions of discrete ill-posed problems as the least squares problem using standard methods of numeric linear algebra such as LU, Cholesky, QR factorizations are unstable. It means that small perturbations in the input data lead to large perturbations in the solution.

Ill-posed problems abound in many areas of science and engineering. Typical examples of discrete ill-posed problems arise from discretization of continuous ill-posed problems, such as Fredholm integral equations of the first kind. Important problems of spectrometry [Zabulonov, Korostil, & Revunova, 2006], gravimetry [Bulakh, 2006], magnitometry [Strakhov, 2008], electrical prospecting [Khmelevskii & Bondarenko, 1999], etc. have the properties of discrete ill-posed problems.

To overcome the instability and to improve the accuracy of solutions to discrete ill-posed problems, regularization methods have been proposed [Hansen, 1998; Tikhonov & Arsenin, 1977; Morozov, 1984; Engl, Hanke, & Neubaer, 2000]. Regularization imposes some constraints on the desired solution that stabilizes the problem and

leads to meaningful and stable solution. For example, Tikhonov regularization [Tikhonov & Arsenin, 1977; Hansen, 1998] penalizes solutions with large l_2-norms.

The drawbacks inherent in the methods of solving discrete ill-posed problems based on Tikhonov regularization include their high computational complexity and the difficulty of selecting the proper regularization parameter (penalty weight) which influences the solution stability. Therefore, alternative approaches are required for solving discrete ill-posed problems that would have the accuracy comparable to Tikhonov regularization at lower computational costs.

We develop such an approach using the ideas of our previous work on distributed representations and random projections [Misuno, Rachkovskij, & Slipchenko, 2005; Revunova & Rachkovskij, 2009]. Recently, researchers working in the area of numeric linear algebra applied similar ideas to get fast randomized algorithms for the least squares problem, matrix factorization, principal component analysis, etc. It is therefore of interest to study those techniques and apply them to discrete ill-posed inverse problems.

This paper is structured as follows. In the first three sections, we provide a brief survey of linear least squares and discrete ill-posed problems, approximate matrix decompositions with randomized algorithms, and randomized least squares approximations. Then, we describe the approach we use to solve discrete ill-posed problems by several methods, including those employing random projections. We provide the bias-variance decomposition of solution error and show experimentally that it has a minimum at some value of the varied smaller dimension of the random projection matrix.. Taking two well-known test examples of discrete ill-posed problems of Carasso and Baart, we obtain experimental results of their solution using both Tikhonov regularization and pseudo-inverse after a random projection. The comparison shows that the minimal error of the pseudo-inverse solution after a random projection is at the level of the standard Tikhonov regularization, but the solution is obtained faster, especially at the larger noise values. The final section provides conclusions and directions of future work.

Linear least squares and discrete ill-posed problems

Many applications in mathematics, physics, data analysis, etc. require finding an approximate solution to a system of linear equations that has no exact solution. For example, (1) with $\mathbf{A} \in \Re^{m \times n}$ for $m>n$ or $m<n$ or $\min(m,n) \neq \text{rank}(\mathbf{A})$ generally does not have $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{b}$. The method of least squares selects $\mathbf{x}$ that minimizes the sum of squares of the elements of the residual vector by solving the optimization problem

$$\mathbf{x'} = \text{argmin}_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}||_2. \quad (2)$$

It is well-known that the minimum-length vector among those satisfying (2) may be computed based on the Moore-Penrose pseudo-inverse of the matrix $\mathbf{A}$ [Demmel, 1997]:

$$\mathbf{x'} = \mathbf{A}^{+}\mathbf{b}. \quad (3)$$

In particular, the traditional solution can be obtained by taking the derivative of $||\mathbf{Ax} - \mathbf{b}||^2 = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b})$ with respect to $\mathbf{x}$ and setting it equal to zero. This gives the system of the so-called normal equations

$$\mathbf{A}^T\mathbf{Ax'} = \mathbf{A}^T\mathbf{b} \quad (4)$$

that should be solved to provide the minimizing vector $\mathbf{x'}$. This requires the residual vector $(\mathbf{Ax'}-\mathbf{b})$ to be orthogonal to the column space of $\mathbf{A}$, i.e., $(\mathbf{Ax'} - \mathbf{b})^T \mathbf{A} = 0$. If the matrix $\mathbf{A}^T\mathbf{A}$ has full numerical rank and is well-conditioned, (4) can be solved using the Cholesky decomposition $\mathbf{A}^T\mathbf{A} = \mathbf{R}^T\mathbf{R}$, giving $\mathbf{R}^T\mathbf{Rx'} = \mathbf{A}^T\mathbf{b}$, where $\mathbf{R}$ is an upper triangular matrix.

In case **A** is rank-deficient or ill-conditioned, solution by a QR decomposition can be used. Here **A**=**QR** is obtained, where **Q** is an orthogonal matrix and **R** is an upper triangular matrix. Substituting it to (4), one obtains the system $\mathbf{Rx'} = \mathbf{Q}^T\mathbf{b}$ that can be immediately solved sequentially, by the backward substitution, since **R** is triangular.

The Singular Value Decomposition (SVD) can also be used. SVD represents **A** as

$$\mathbf{A} = \mathbf{USV}^T, \tag{5}$$

where **U** is the matrix of left singular vectors with orthonormal columns, and **V** is the matrix of right singular vectors with orthonormal columns, **S** is the diagonal matrix of singular values.

The Moore-Penrose generalized inverse, or pseudo-inverse, of **A** may be expressed in terms of SVD as

$$\mathbf{A}^+ = \mathbf{VS}^{-1}\mathbf{U}^T. \tag{6}$$

Then the solution **x'** is obtained by (3).

Known techniques find a solution to (2) in $O(mn^2)$ time. The SVD method is the most computationally intensive among those mentioned above, but it can be useful if **A** is very ill-conditioned. Let us consider this in more detail. Denote the i-th singular value of **A** as $\sigma_i(\mathbf{A})$, $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq ... \geq 0$, and the maximum and minimum singular value of **A** as $\sigma_{max}(\mathbf{A}) = \sigma_1(\mathbf{A})$ and $\sigma_{min}(\mathbf{A})$. The condition number of **A** is $\text{cond}(\mathbf{A}) = \sigma_{max}(\mathbf{A})/\sigma_{min}(\mathbf{A})$. When cond(**A**) is large, the matrix **A** is ill-conditioned, implying that the solution is potentially unstable and inaccurate. An indication of instability is the fact that small changes in the vector **b** (e.g. due to the noise) cause large changes in the solution **x'**, and the solution error is typically large, especially when the noise level increases. Actually, for large cond(**A**) the inverse of singular values in $\mathbf{S}^{-1}$ become very large, and so the values of **x'** components, due to (3) and (6).

If there is a sharp gap in the spectrum of singular values, and the singular values after the gap are very small, they may be considered as noisy ones and are eliminated in some implementations by thresholding to overcome the instability. However, this does not work for discrete ill-posed problems, since, as mentioned in the Introduction, they does not have the gap in their singular values, and so their numerical rank is ill-determined and the choice of an appropriate numerical rank is not easy.

The classical method of solving discrete ill-posed problems is Tikhonov regularization [Tikhonov & Arsenin, 1977; Hansen, 1998]. The standard form of Tikhonov regularization problem is formulated as follows

$$\mathbf{x}_{reg} = \text{argmin}_{\mathbf{x}} \ (||\mathbf{Ax}-\mathbf{b}||_2+\lambda||\mathbf{x}||_2), \tag{7}$$

where λ is the regularization parameter.

Solution of (7) can be obtained by the method of filtered SVD [Hansen, 1998]:

$$\mathbf{x}_{reg} = \mathbf{V}\ \text{diag}\ (f_i / \sigma_i)\ \mathbf{U}^T\mathbf{y}, \tag{8}$$

where $f_i = \sigma^2_i /(\sigma^2_i + \lambda^2)$ are filter factors.

One problem here is using the SVD of **A**, since it is a computationally expensive decomposition method. Another problem, that may appear even more important, is selecting the proper regularization parameter λ.

A number of methods for selecting the regularization parameter have been proposed [Engl, Hanke, & Neubaer, 2000]. The L-curve method [Hansen & O'Leary, 1993] makes a plot of $||\mathbf{x}_{reg}||_2$ vs $||\mathbf{Ax}_{reg} - \mathbf{b}||_2$ for all valid regularization parameters. For discrete ill-posed problems the L-curve, when plotted in log-log scale, often has a characteristic L-shape appearance (hence its name). A distinct corner separates the vertical and the horizontal parts of the curve. The regularization parameter not far from the corner is selected as optimal.

The discrepancy principle [Morozov, 1984] chooses the regularization parameters such that the residual norm for the regularized solution satisfies $||\mathbf{A}\mathbf{x}_{reg} - \mathbf{b}||_2 = ||\mathbf{e}||_2$, where **e** in the norm of perturbation of the right-hand side, and therefore requires an estimation of $||\mathbf{e}||_2$.

The generalized cross-validation [Wahba, 1990] method is based on the idea that an arbitrary element b_i of the right-hand side **b** can be predicted by the corresponding regularized solution, and the choice of regularization parameter should be independent of an orthogonal transformation of **b**. This leads to choosing the regularization parameter that minimizes $||\mathbf{A}\mathbf{x}_{reg} - \mathbf{b}||^2 / D^2$, where D is a squared effective number of degrees of freedom (which is not necessarily an integer) that can be calculated as $D = m - \Sigma_i f_i$. Here the errors of **b** are considered as uncorrelated zero-mean random variables with a common variance, i.e., white noise.

However, those methods do not always produce stable results. So, at the wrong values of the regularization parameter the error of solution may be substantial.

Now, let us consider some recent randomized approaches whose ideas, in our opinion, may be useful for solving discrete ill-posed problems. Though randomization is aimed at faster computations, we are also interested in its employing for stabilizing the solution of discrete ill-posed problems.

Approximate matrix decompositions with randomized algorithms

As follows from the previous sections, matrix decompositions play a major role in solutions of least squares problems and are used in many numerical computer science applications. Let us briefly consider recent approaches to approximate matrix decompositions with randomized algorithms (for an extended review, see [Halko, Tropp & Martinsson, 2009]).

The main motivations of research in this area are as follows. Today, data sets represented by matrices are very large, so that classical algorithms are not always well suited for their processing. Also, traditional algorithms were designed to produce highly accurate matrix decompositions, but it seems that this is not always justified, since the data in large matrices are often inherently imprecise. Other motivations include overcoming the data transfer bottleneck that may dominate the computational costs by getting algorithms with fewer passes over the data, and exploiting the power of novel processor architectures optimized for specific vector operations.

It should be noted that randomized algorithms are often more accurate and robust than deterministic ones, as their error bounds show. And, though their results are probabilistic, the probability of failure may be set by parameter choice to be negligible (e.g., less than 10^{-15}) while preserving the computational power advantages.

Randomization may be accomplished by random sampling or random projection or their combination. Sampling is considered here as getting some random subset of matrix, such as individual entries (components, cells) or columns and rows, according to some probability distribution. Projecting is a random linear mapping (embedding), usually implemented using multiplication by a random matrix whose components are generated according to some probability distribution. Both sampling and projecting reduce the dimensionality of the initial task and so decrease the computational complexity of subsequent processing.

A low-rank approximation of a given matrix is expressed as

$$\mathbf{A}=\mathbf{BC},\ \mathbf{A}\in\Re^{m\times n},\ \mathbf{B}\in\Re^{m\times k},\ \mathbf{C}\in\Re^{k\times n}. \tag{9}$$

The inner dimension k is sometimes called the numerical rank of the matrix.

The fixed-precision approximation problem is formulated as finding **A'** so that

$$\min||\mathbf{A} - \mathbf{A}'||_2 < \varepsilon, \tag{10}$$

whereas the fixed-rank approximation problem is finding **A'** given its target rank *k* so that

$$\min_{\text{rank}(\mathbf{A}')\le k} ||\mathbf{A} - \mathbf{A}'||_2. \quad (11)$$

The task of a low-rank approximation to a given matrix can be efficiently computed using randomization by the following two stages.

Stage I (the randomization stage) is to construct a low-dimensional randomized "reduced representation" of the input matrix **A** that captures its action. This includes the following steps.

Step 1. Make a randomized procedure to get a reduced representation of **A**.

This procedure is represented as an operator (matrix) **X**. The random matrix **X** is carefully constructed, often as a product of several matrices, to ensure the required computational complexity and accuracy of the final approximation.

Step 2. Apply **X** to **A** by forming the matrix product **Y** = **AX**.

Here **Y** contains the obtained reduced representation of **A**. Explicit dimensionality and/or rank of **Y** is less than that of **A**, thus ensuring a lower computational complexity of the second stage.

Stage II (the deterministic stage) is to postprocess the result of Stage I to compute a final approximation, typically with well-established deterministic methods from numerical linear algebra, such as a standard factorization (QR, SVD, etc.). This step may require another look at the matrix **A**.

In some procedures those stages may be distinguished not so clearly, but the main idea is preserved: use some random transformation of **A** to accelerate computing of the final result.

In the matrix approximation framework, Stage I generally dominates the cost of Stage II. Within Stage I, the computational bottleneck is usually the matrix-matrix product **Y** = **AX** in Step 2. However, for some specific constructs of **X** and/or computational architectures this multiplication can be implemented very efficiently. Anyway, its cost is usually much lower that the cost of **A** factorization by the traditional methods – otherwise, randomization should not be used.

As an example, let us consider the task of computing an approximate SVD of $\mathbf{A}\in\Re^{m\times n}$ with a target numerical rank *k* by the Randomized PCA Algorithm, as considered in [Halko, Tropp & Martinsson, 2009] after [Liberty et al., 2007; Sarlos, 2006].

The randomizing Stage I looks as follows:

I.1. As a randomized procedure, generate a projector matrix $\mathbf{G}\in\Re^{n\times(k+p)}$. The elements of **G** are the realizations of a Gaussian random variable with zero mean and unit variance; *p* is usually a small integer. (12)

I.2. Apply **G** to **A** to obtain **Y** = **AG**.

The classical deterministic stage (Stage 2) includes the following steps:

II.1 Construct a matrix **Q** whose columns form an orthonormal basis for the range of **Y**: **Y**=**QR**.

II.2 Form **B** = **QA.**

II.3 Compute an SVD of the small matrix: **B** = **U' S V**. (13)

II.4 Set **U** = **QU'**.

In order to reduce the error bound of this approximation, this technique were elaborated in [Rokhlin, Szlam, & Tygert, 2009]. There, $p=k$, and I.2 is changed to $\mathbf{Y}=(\mathbf{AA}^{\mathrm{T}})^{q}\,\mathbf{AG}$, where q are the steps of a power iteration, q = 1, 2 is usually sufficient in practice.

The randomized PCA algorithm we have just considered is an instantiation of the **dimension reduction** (or **projective**) **approach** to matrix approximation. The rationale behind it is that since the rows of a low-rank matrix are linearly dependent, they can be embedded into a low-dimensional space preserving their geometric properties. A random linear map **G** provides an efficient way to perform this embedding: collecting random samples from the column space of the matrix is equivalent to reducing the dimension of the rows. Adaptation to **A** is not required to obtain the sampling distribution. Taking samples requires substantial computations, but can be optimized. Then, the samples are orthogonalized as preparation for constructing various matrix approximations.

The work on randomized matrix approximations has been substantially influenced by the field of random embeddings. [Johnson & Lindenstrauss, 1984] showed that the pairwise Euclidean distances between the points in the input space are approximately preserved when the points are mapped to a Euclidean space of a much smaller dimension using a random Gaussian projector. This and subsequent work suggested that some computational problems of a geometric nature may be solved in a more computationally efficient manner by first translating them into a lower-dimensional space. This was used in some applications that require a fast estimation of vector similarity. [Papadimitriou et al., 2000] were the first who applied this approach to linear algebra in context of "latent semantic indexing" based on SVD.

Some researchers studied the issues of simplifying the mapping itself and the cost of its applying. [Achlioptas, 2003] demonstrated that discrete random matrices with {–1,0,1} components perform nearly as well as a Gaussian matrix. [Li, Hastie, & Church, 2006] considered very sparse (and therefore potentially fast) discrete random matrices, however, as we know, so far their work has not been applied to numerical linear algebra. On the other hand, the fast Johnson-Lindenstrauss transform [Ailon & Chazelle, 2006] which combines the speed of the Fast Fourier Transform with the embedding properties of a Gaussian matrix (further developed in [Ailon & Liberty, 2008; Liberty, Ailon, & Singer, 2008]) has got substantial interest in the area under consideration. In particular, [Sarlos, 2006] applied these techniques to matrix approximation, which has led to some of the fastest algorithms available today [Liberty et al., 2007; Woolfe et al., 2008].

[Woolfe et al., 2008] propose a fast randomized algorithm for the approximation of matrices by constructing and applying a structured random matrix $\mathbf{R}\in\Re^{l\times m}$ ($l > k$) to each column of **A**, $\mathbf{Y}=\mathbf{RA}$. Here **R** is obtained with the subsampled random Fourier transform. The structure of **R** allows one to apply it to an arbitrary m-vector rapidly, at a cost proportional to $m\log(l)$. So, the resulting procedure can construct a rank-k approximation from the entries of **A** at a cost proportional to $mn\log(k)+l^2(m+n)$.

The accuracy bound of the algorithm guarantees that the spectral norm $\|\mathbf{A}-\mathbf{A}'\|$ is of the same order as $(\max\{m,n\})^{1/2}\sigma_{k+1}$ of **A**. Note, that in order to compute a similarly accurate rank-k approximation, the standard approach to approximate SVD by computing the full SVD and truncating it requires $O(mn\ \min\{m;n\})$ floating-point operations. A more efficient deterministic scheme using a partial QR factorization and factor postprocessing takes $O(kmn)$ operations. So, the classical decomposition algorithms require more operations than the present one [Woolfe et al, 2008]. In practice, the algorithm runs faster than the classical algorithms, even when k is quite small or large.

Another approach to randomized matrix approximation may be considered as **the sampling approach**. Its simplest manifestation is matrix sparcification that randomly preserves a fraction of non-zero entries of **A** with the aim to reduce storage or/and to accelerate matrix computations. Another related instance of this approach stems from the methods that build a matrix approximation from a submatrix and computed coefficient matrices.

Approximations using a subset of columns are known as the interpolative decomposition; those using a subset of rows and a subset of columns are known as the CUR decomposition.

The interpolative decomposition is based on the idea that a small set of columns describes most of the action of a numerically low-rank matrix [Ruston, 1964]. A number of works develop randomized algorithms for this class of matrix approximations. These methods first compute a sampling probability for each column using their leverage scores that reflect the relative importance of the columns to the action of the matrix. Columns are then selected randomly according to this distribution, and post processed to provide the final matrix approximation. Such a column sampling can also be viewed as an adaptive form of dimension reduction. Realization of this method by [Drineas, Kannan, & Mahoney, 2006] showed that, given a target rank k and a parameter $\varepsilon>0$, it produces the approximation matrix **B** for which

$$||\mathbf{A} - \mathbf{B}||_F \leq || \mathbf{A} - \mathbf{A}_{(k)} ||_F + \varepsilon ||\mathbf{A}||_F, \quad (14)$$

where $||\cdot||_F$ denotes the Frobenius norm, $\mathbf{A}_{(k)}$ is a best rank-k approximation of **A**. The same column sampling method also yields spectral-norm error bounds [Rudelson & Vershynin, 2007].

[Deshpande et al., 2006; Deshpande & Vempala, 2006] and [Har-Peled, 2006] demonstrated that the error in the column sampling approach can be improved by an iterative k-pass algorithm for which

$$||\mathbf{A} - \mathbf{B}||_F \leq (1+\varepsilon) || \mathbf{A} - \mathbf{A}_{(k)} ||_F. \quad (15)$$

Subsequent work [Boutsidis, Mahoney, & Drineas, 2008; Boutsidis, Drineas, & Mahoney, 2009] showed that post processing the sampled columns with a rank-revealing QR algorithm can reduce the number of output columns required while improving the classical existential error bound [Ruston, 1964].

Drineas et al. have developed randomized techniques for computing CUR decompositions, which express $\mathbf{A} \approx \mathbf{CUR}$, where **C** and **R** are small column and row submatrices of **A,** and **U** is a small linkage matrix. These methods identify columns (rows) that approximate the range (corange) of the matrix; the linkage matrix is then computed by solving a small least-squares problem. [Drineas, Kannan, & Mahoney, 2006a] developed a randomized CUR algorithm with controlled absolute error, and [Drineas, Mahoney, & Muthukrishnan, 2008] developed a relative error algorithm.

Randomized Least Squares Approximations

In this section we consider recent work aimed at a rapid solution of the least squares approximation problem. Though the main ideas were considered in the previous section, in this area they are immediately applied to this very problem, and the error bounds and complexity are investigated directly for this case.

As before, randomization is viewed as preconditioning the input matrix **A**, and now also the target vector **b** with a carefully-constructed data-independent random matrix **X** (often constructed from several other matrices). Thus, the least squares approximation problem (2) is replaced with

$$\mathbf{x}' = \text{argmin}_{\mathbf{x}} ||\mathbf{Xb} - \mathbf{XAx}||_2. \quad (16)$$

Solution to (16) can be explicitly computed using some traditional deterministic algorithm, e.g., SVD, to compute the generalized inverse

$$\mathbf{x}' = (\mathbf{XA})^{+} \mathbf{Xb}. \quad (17)$$

Alternatively, standard iterative methods could be used, such as the Conjugate Gradient Normal Residual method. It can produce an ε-approximation to the optimal solution of (16) in $O(\text{cond}(\mathbf{XA})\, m \log(1/\varepsilon))$ time, where cond(**XA**) is the condition number of **XA**, and r is the number of rows of **XA**.

The idea to apply randomization to the least squares problem using sampling was proposed in [Drineas, Mahoney, & Muthukrishnan, 2006]. [Sarlos, 2006] pioneered applying a random projection matrix to this problem. The error bounds and speed for both approaches were improved in [Drineas et al., 2007] for the over constrained ($m>>n$) least squares approximation problems.

Their random sampling algorithm starts by preprocessing the matrix **A** and the right hand side vector **b** with a randomized Hadamard transform **H**. It then constructs a smaller problem by uniformly randomly sampling a small number r of constraints from the preprocessed problem. So, here **X** = **SH**, where **S** is a matrix that represents the sampling operation. Then, this algorithm solves the least squares problem on just those sampled constraints.

In a similar manner, the second algorithm also initially transforms the input **A** and **b** by **H**. Then it multiplies the result by a sparse projection matrix **T** of $O(n/\varepsilon) \times m$. So, here **X** = **TH**. Finally, the second algorithm solves the least squares problem on just those $O(n/\varepsilon)$ coordinates to obtain n-vector **x'**.

In both cases, the solution to the smaller problem provides a relative-error ε-approximation to the exact solution, while making no assumptions on the input data. And if m is sufficiently larger than n, the approximate solution can be computed in $o(mn^2)$ time. Traditional methods (of Gauss, etc.) find a solution in $O(mn^2)$ time.

In [Rokhlin & Tygert, 2008] similar techniques were used to analyze theoretically and evaluate empirically random-projection based algorithms for over determined least-squares problems. **X** implements the subsampled randomized Fourier transform. Their procedure computes an n-vector **x** such that **x** minimizes the Euclidean norm $\mathbf{Ax}-\mathbf{b}$ to a relative precision ε. The algorithm typically requires $O((\log(n) + \log(1/\varepsilon))mn + n^3)$ floating-point operations, whereas the earlier algorithms involve costs proportional to $1/\varepsilon$. This cost is less than $O(mn^2)$ required by the classical schemes based on QR-decompositions or bidiagonalization. Moreover, the methods extend to underdetermined ($m<n$) least-squares [Tygert, 2009].

Solving discrete ill-posed problems using random projections

Now let us consider discrete ill-posed problem in the form:

$$\mathbf{\Phi x} = \mathbf{y}. \tag{18}$$

Here we have the matrix $\mathbf{\Phi} \in \Re^{N\times N}$ with singular values σ_i gradually decaying zero and large condition number. The vector $\mathbf{y} \in \Re^N$ is distorted by the additive noise $\boldsymbol{\varepsilon} \in \Re^N$: $\mathbf{y}=\mathbf{y}_0+\boldsymbol{\varepsilon}$.

Let us use the randomized algorithms (described in the survey sections above) not only to accelerate, but also to stabilize the solution **x'** of the ill-posed problem, as follows. Multiply both sides of (18) by the matrix $\mathbf{\Omega} \in \Re^{K\times N}$, $K \le N$, whose elements are realizations of a normal random variable with zero mean and unit variance. The number of columns N of matrix Ω is determined by the dimension of the matrix **Φ**, the number of rows K is a priori unknown since the numerical rank of **Φ** is ill-determined and the required numerical rank of approximation is unknown. We obtain

$$\mathbf{\Omega\Phi\, x} = \mathbf{\Omega y}, \text{ where } \mathbf{\Omega\Phi} \in \Re^{K\times N}, \mathbf{\Omega y} \in \Re^K. \tag{19}$$

Then the least-squares problem is

$$\mathbf{x}_{Pr} = \text{argmin}_{\mathbf{x}} \|\mathbf{\Omega\Phi\, x} - \mathbf{\Omega y}\|_2. \tag{20}$$

Signal reconstruction based on pseudo-inverse is obtained as

$$\mathbf{x}_{pinPr} = (\mathbf{\Omega\Phi})^{+}\,\mathbf{\Omega y}. \tag{21}$$

Signal reconstruction by Tikhonov regularization is obtained as

$$\mathbf{x}_{regPr} = \mathrm{argmin}_{\mathbf{x}}\,(\,||\mathbf{\Omega\Phi x}-\mathbf{\Omega y}||_2+\lambda||\mathbf{x}||_2\,). \tag{22}$$

The accuracy of solving the inverse problem will be evaluated using the error *d* of recovery of the true signal **x**0:

$$e = ||\mathbf{x}0 - \mathbf{x'}||_2 = ||\mathbf{e}||_2, \tag{23}$$

where **x'** is the reconstructed signal vector, **e** is the solution error vector.

The error vector **e** is represented [Vogel, 2002; Goldenshluger & Pereverzev, 2000] as the sum of bias and variance. Let us calculate them for our task as follows. Denote P the operator that transforms y to x': x' = Py. Then, taking into account that $\mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}$ and $\mathbf{y}_0 = \mathbf{\Phi x}0$, the expression for x' can be represented as:

$$\mathbf{x'} = \mathbf{P}\,(\mathbf{y}_0 + \boldsymbol{\varepsilon}) = \mathbf{P}\,\mathbf{y}_0 + \mathbf{P}\,\boldsymbol{\varepsilon} = \mathbf{P}\,\mathbf{\Phi x}0 + \mathbf{P}\,\boldsymbol{\varepsilon}. \tag{24}$$

Using the expression for **x'**, we obtain the expression for **e**:

$$\mathbf{e} = \mathbf{x'} - \mathbf{x}0 = \mathbf{P}\,\mathbf{\Phi x}0 - \mathbf{x}0 + \mathbf{P}\,\boldsymbol{\varepsilon} = (\mathbf{P}\,\mathbf{\Phi} - \mathbf{I})\mathbf{x}0 + \mathbf{P}\,\boldsymbol{\varepsilon}. \tag{25}$$

Thus,

$$\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2, \text{ где } \mathbf{e}_1,\ \mathbf{e}_2 \in \Re^N,\ \mathbf{e}_1 = (\mathbf{P}\,\mathbf{\Phi} - \mathbf{I})\mathbf{x}0,\ \mathbf{e}_2 = \mathbf{P}\,\boldsymbol{\varepsilon}. \tag{26}$$

$\mathbf{e}_1$ is called bias, $\mathbf{e}_2$ is called variance [Vogel, 2002; Goldenshluger & Pereverzev, 2000].

To obtain the solution **without projection**, we use the following methods.

The pseudo-inverse solution based on SVD is actually computed as:

$$\mathbf{x}_{pin} = \mathbf{\Phi}^{+}\,\mathbf{y},\ \mathbf{P}_{pin} = \mathbf{\Phi}^{+} = \mathbf{V}\ \mathrm{diag}\,(\varphi_i / \sigma_i)\ \mathbf{U}^{\mathrm{T}},\ \text{iff } \sigma_i > tresh\ \varphi_i = 1,\ \text{otherwise } \varphi_i = 0.$$
$$tresh = max(K,N)\ \mathrm{eps}(max(\sigma_i)), \tag{27}$$

where **U**, **V, S** are obtained by SVD of $\mathbf{\Phi} = \mathbf{USV}^{\mathrm{T}}$; σ_i = diag **S** are singular values, the elements of a diagonal matrix **S**; floating-point relative accuracy eps(*z*) is the positive distance from abs(*z*) to the next larger in magnitude floating point number of the same precision as z.

Here the bias-variance decomposition is

$$\mathbf{e}_{1pin} = (\mathbf{P}_{pin}\,\mathbf{\Phi} - \mathbf{I})\mathbf{x}0,\ \mathbf{e}_{2pin} = \mathbf{P}_{pin}\,\boldsymbol{\varepsilon}. \tag{28}$$

For the solution based on Tikhonov regularization (22)

$$\mathbf{P}_{reg} = \mathbf{V}\ \mathrm{diag}\,(f_i / \sigma_i)\ \mathbf{U}^{\mathrm{T}}. \tag{29}$$

$$\mathbf{e}_{reg} = \mathbf{x}_{reg} - \mathbf{x}0 = \mathbf{P}_{reg}(\mathbf{\Phi x}0 + \boldsymbol{\varepsilon}) - \mathbf{x}0 = (\mathbf{P}_{reg}\,\mathbf{\Phi} - \mathbf{I})\,\mathbf{x}0 + \mathbf{P}_{reg}\,\boldsymbol{\varepsilon}; \tag{30}$$

$$\mathbf{e}_{1reg} = (\mathbf{P}_{reg}\,\mathbf{\Phi} - \mathbf{I})\,\mathbf{x}0,\ \mathbf{e}_{2reg} = \mathbf{P}_{reg}\,\boldsymbol{\varepsilon}. \tag{31}$$

After the projection, the bias-variance decomposition of the solution error vector takes the following forms.

By analogy to (25), we can write:

$$\mathbf{e}_{Pr} = \mathbf{x}_{Pr} - \mathbf{x}0 = \mathbf{P}_{Pr}\,\mathbf{\Omega}\,\mathbf{y} - \mathbf{x}0 = \mathbf{P}_{Pr}\,\mathbf{\Omega}\,(\mathbf{\Phi x}0 + \boldsymbol{\varepsilon}) - \mathbf{x}0 = (\mathbf{P}_{Pr}\,\mathbf{\Omega\Phi} - \mathbf{I})\,\mathbf{x}0 + \mathbf{P}_{Pr}\,\mathbf{\Omega}\boldsymbol{\varepsilon}, \tag{32}$$

where $\mathbf{x}_{Pr}$ is the solution after projection, $\mathbf{P}_{Pr}$ is the operator that transforms $\mathbf{\Omega}$y to $\mathbf{x}_{Pr}$: $\mathbf{x}_{Pr} = \mathbf{P}_{Pr}\,\mathbf{\Omega y}$.

For the pseudo-inverse solution after the projection:

$$\mathbf{P}_{pinPr} = \mathbf{C}\ \mathrm{diag}\,(\varphi_i / s_i)\ \mathbf{B}^{\mathrm{T}}, \tag{33}$$

where **B**, **C**, **Σ** are the SVD result for the matrix $\mathbf{\Omega\Phi} = \mathbf{B\,\Sigma\,C}^{\mathrm{T}}$, $s_i = \mathrm{diag}\,\mathbf{\Sigma}$ are singular values, the elements of the diagonal matrix **Σ**, s_i are thresholded by φ_i as in (27). The error terms are

$$\mathbf{e}_{1\mathrm{pinPr}} = (\mathbf{P}_{\mathrm{pinPr}}\,\mathbf{\Omega\Phi} - \mathbf{I})\mathbf{x}0,\ \mathbf{e}_{2\mathrm{pinPr}} = \mathbf{P}_{\mathrm{pinPr}}\,\mathbf{\Omega\varepsilon};\ \mathbf{e}_{1\mathrm{pinPr}}+\mathbf{e}_{2\mathrm{pinPr}} = \mathbf{e}_{\mathrm{pinPr}}, \tag{34}$$

where $\mathbf{e}_{\mathrm{pinPr}}$ is the solution error vector for the pseudo-inverse using the random projection.

For Tikhonov regularization after the projection:

$$\mathbf{P}_{\mathrm{regPr}} = \mathbf{C}\,\mathrm{diag}(\varphi_i / s_i)\,\mathbf{B}^{\mathrm{T}}, \tag{35}$$

where $\varphi_i = s^2_i /(s^2_i + \lambda^2)$ are filter multipliers.

$$\mathbf{e}_{1\mathrm{regPr}} = (\mathbf{P}_{\mathrm{regPr}}\,\mathbf{\Omega\Phi} - \mathbf{I})\mathbf{x}0,\ \mathbf{e}_{2\mathrm{regPr}} = \mathbf{P}_{\mathrm{regPr}}\,\mathbf{\Omega\varepsilon};\ \mathbf{e}_{1\mathrm{regPr}}+\mathbf{e}_{2\mathrm{regPr}} = \mathbf{e}_{\mathrm{regPr}}, \tag{36}$$

where $\mathbf{e}_{\mathrm{regPr}}$ is the solution error vector for Tikhonov regularization using the random projection.

For the solution based on the method of approximate matrix factorization (AMF) (12)-(13).

$$\mathbf{P}_{\mathrm{amfPr}} = (\mathbf{\Phi'})^{+} = \mathbf{V}\,\mathrm{diag}(\varphi_i / z_i)\,\mathbf{U}^{\mathrm{T}}, \tag{37}$$

where $\mathbf{\Phi'}$ is the result of **Φ** approximation after projection using $\mathbf{\Omega} \in \Re^{N\times(K+2)}$, **V** is the matrix obtained by (13.II.3), **U** is obtained by (13.II.4), the singular values z_i are thresholded by φ_i as in (27).

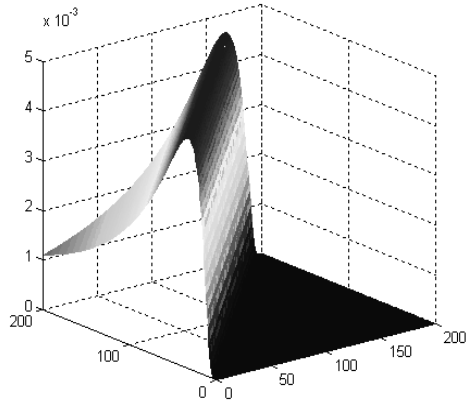

Fig.1. The matrix **Φ** of the Carasso problem.

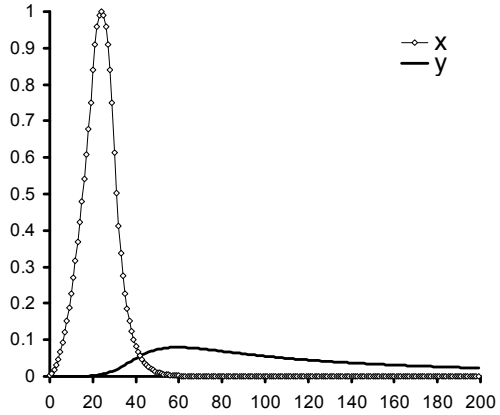

Fig. 2. The signal vector **x** and the right hand side **y** of the Carasso problem.

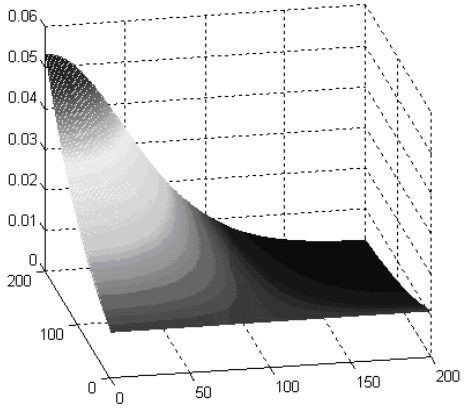

Fig. 3. The matrix **Φ** of the Baart problem.

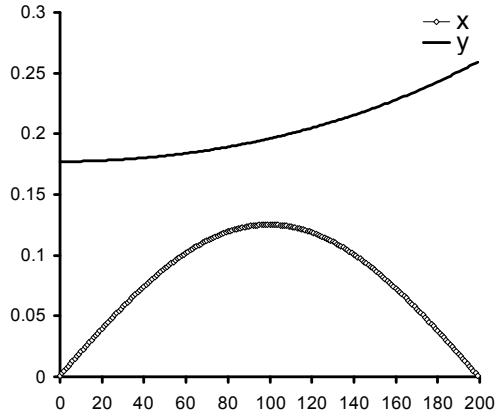

Fig.4. The signal vector **x** and the right hand side **y** of the Baart problem.

The error terms are

$$\mathbf{e}_{1amfPr}=(\mathbf{P}_{amfPr}-\mathbf{I})\mathbf{x}0,\ \mathbf{e}_{2mfPr}=\mathbf{P}_{amfPr}\,\boldsymbol{\varepsilon};\ \mathbf{e}_{1amfPr}+\mathbf{e}_{2amfPr}=\mathbf{e}_{amfPr}, \tag{38}$$

where $\mathbf{e}_{amfPr}$ is the solution error vector for the approximate matrix factorization using the random projection (12)-(13).

Experiments

Let us conduct an experimental investigation using two well-known discrete ill-posed problem.

The problem of Carasso [Carasso, 1982] seeks to reconstruct the time profile of a heat source by monitoring the temperature at a fixed distance away. The inverse heat equation here is a Volterra integral equation of the first kind with [0,1] as integration interval. The kernel is

$$K(s,t)=k(s-t),\ k(t)=t^{3/2}/2\pi^{1/2}\exp(-1/4t). \tag{39}$$

The integral equation is discretized by the quadrature method using simple collocation and the midpoint rule with N points (e.g., [Carasso, 1982]). An exact solution $\mathbf{x}_0$ is constructed, and then $\mathbf{y}_0$ is produced as $\mathbf{y}_0=\boldsymbol{\Phi}\mathbf{x}_0$.

The problem of Baart [Baart, 1982] is discretization of an artificial Fredholm integral equation of the first kind with kernel K and right-hand side h given by

$$K(s,t)=\exp(s\cos(t)),\ h(s)=2\sin(s)/s;\ \ s\in[0,\pi/2],\ t\in[0,\pi]. \tag{40}$$

Discretization is done by the Galerkin method with orthonormal basis functions. The known solution is also represented as a vector $\mathbf{x}_0$. In this problem, $\boldsymbol{\Phi}\mathbf{x}_0$ is not exactly equal to $\mathbf{y}_0$.

In both problems, the matrix $\boldsymbol{\Phi}$ has dimensionality of 200×200. The right-hand side y0 is distorted by an additive noise with the uniform distribution and various amplitudes. The random projection matrix $\boldsymbol{\Omega}\in\Re^{K\times N}$, N=200, $K\le N$ is the Gaussian random matrix with the entries of zero mean and unit variance.

Let us consider how the random projection changes the singular value spectrum. In Fig. 5, the singular values of $\boldsymbol{\Phi}$, $\boldsymbol{\Omega\Phi}$ (K=30), $\boldsymbol{\Phi}'$ (K=30) are shown for the Carasso problem. Evidently, the numerical rank of $\boldsymbol{\Phi}$ is ill-determined, whereas the rank of $\boldsymbol{\Omega\Phi}$ and $\boldsymbol{\Phi}'$ is equal to 30.

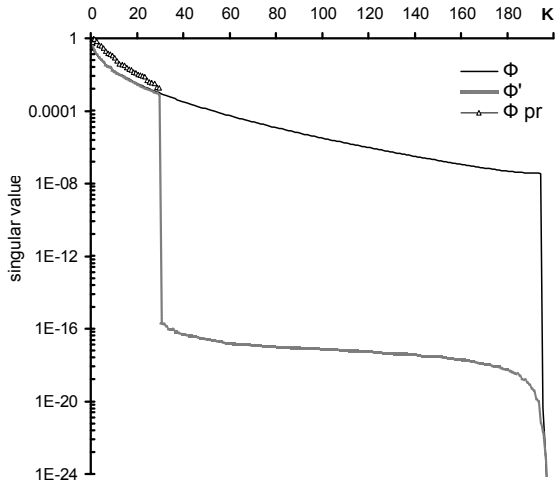

Fig. 5. The singular values of the Carasso problem for $\boldsymbol{\Phi}$ (Φ), $\boldsymbol{\Omega\Phi}_{K=30}$ (Φ pr), $\boldsymbol{\Phi}'_{K=30}$ (Φ').

Fig. 6 shows the dependence of the signal reconstruction error *e* on the row number *K* of the projector matrix $\mathbf{\Omega}(K \times N)$ the Carasso problem. The plots are provided for three solutions that use the projected matrix $\mathbf{\Omega\Phi}$: the pseudo-inverse (20)-(21), the Tikhonov regularization (22) with the selection of λ using the discrepancy principle [Morozov, 1984], and the pseudo-inverse of the approximate matrix $\mathbf{\Phi'}$ obtained by (12)-(13).

Three noise levels are used. For all three solution techniques at some value of *K* the minimum of the error *d* can be observed for all noise levels. When the noise level increases, the minimum position shifts towards the smaller values of *K* and the error value *d* at the minimum increases.

Fig.7 shows the dependences of the two error terms on K: the norm of the bias error term e1 = ||e1||2 and variance e2 = ||e2||2 (37)-(38) obtained for the solution by the approximate matrix factorization (12)-(13). With the increasing K, $||\mathbf{e}_1||$ decreases and $||\mathbf{e}_2||$ increases, so that $||\mathbf{e}_1+\mathbf{e}_2||$ has a minimum. When the noise level increases, position of the error minimum shifts towards the lower values of K. This occurs because the dependence of $||\mathbf{e}_1||$ on K is practically the same for all levels of noise, whereas $||\mathbf{e}_2||$ increases with the increasing noise level.

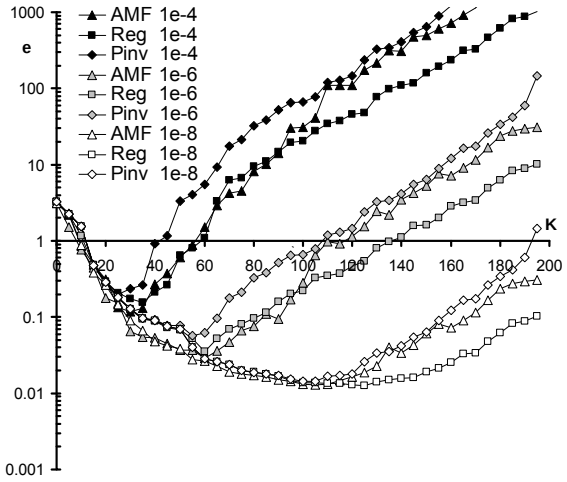

Fig. 6. Dependence of the signal reconstruction error e for the Carasso problem on the row number K of the projection matrix at the noise levels 10–8, 10–6, 10–4. Pinv is pseudo-inverse, Reg is the Tikhonov regularization with the selection of λ using the discrepancy principle, AMF is the approximate matrix factorization.

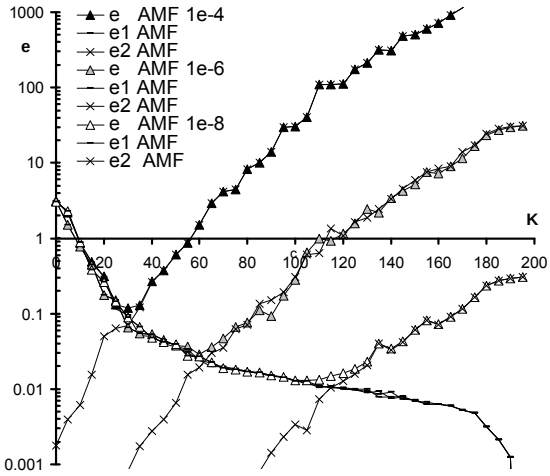

Fig. 7. The dependences of bias e_1, variance e_2, and the total error on *K* at the noise levels 10–8, 10–6, 10–4 for the solution of the Carasso problem by the approximate matrix factorization AMF.

For the Baart problem, the dependences of the signal reconstruction error e (Fig. 8) and the bias e1 and variance e2 terms (Fig. 9) are similar to those of the Carasso problem. Here, bias decreases not so profoundly, probably because for this problem $\mathbf{\Phi x}_0$ is not exactly equal to $\mathbf{y}_0$.

Table 1 shows the error results. For the Carasso problem, for the techniques without projection, the error *e* is given, and for the techniques with projection, the minimum value of errors obtained at some *K* is given. Without projection, the standard pseudo-inverse provides a somewhat acceptable error only at the lowest noise level and cannot be used at the larger noise levels. The error levels for Tikhonov regularization are comparable for all three noise levels, and are at the level the pseudo-inverse error at 10–8.

With projection, the error values at the minimum are comparable and small for all techniques. The pseudo-inverse with projection becomes stable and shows the error values at the level of Tikhonov regularization with projection. The error values are much smaller than those for the Tikhonov regularization without projection.

For the Baart problem, the minimal errors after projection for all techniques are at the level of the Tikhonov regularization without projection, however the error is somewhat smaller for the approximate matrix factorization technique.

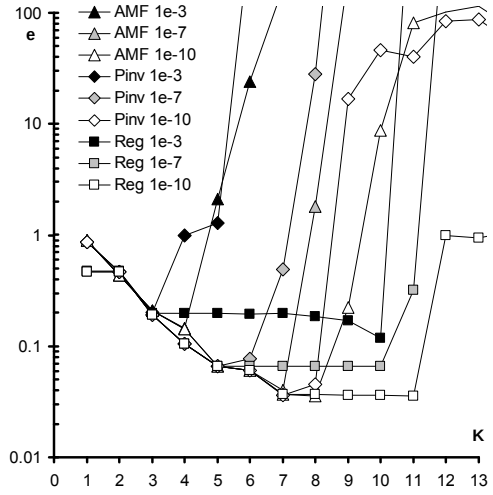

Fig. 8 Dependence of the signal reconstruction error *e* for the Baart problem on the row number *K* of the projection matrix at the noise levels 10^{-10}, 10^{-7}, 10^{-3}. Notations are as in Fig.6.

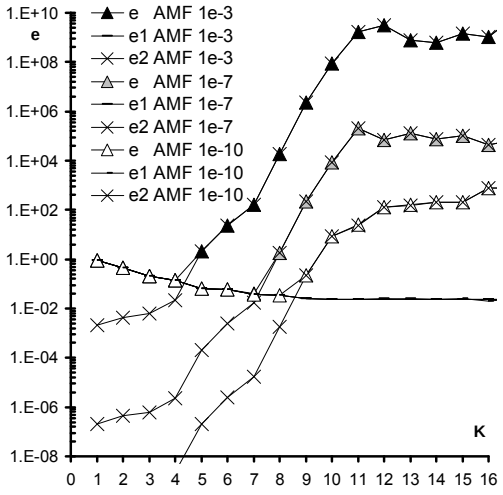

Fig. 9 Dependences of the bias e1 and variance e2 error terms for the Baart problem on *K* for the solution based on the approximate matrix factorization AMF.

Table 1. The solution errors for the Carasso and the Baart problems. e_{pin} and e_{reg} are the errors of the pseudo-inverse and the Tikhonov regularization solutions without projection, e_{pinPr} and e_{regPr} are those with projection, e_{amfPr} is the solution error obtained using the approximate matrix factorization with projection.

The Carasso problem						The Baart problem					
Noise level	Error, solutions w/o projection		Min error, solutions w/ projection			Noise level	Error, solutions w/o projection		Min error, solutions w/ projection		
	e_{pin}	e_{reg}	e_{pinPr}	e_{regPr}	$e_{amf\ Pr}$		e_{pin}	e_{reg}	e_{pinPr}	e_{regPr}	$e_{amf\ Pr}$
10^{-8}	0.308	0.308	0.0136	0.013	0.013	10^{-10}	231	0.037	0.036	0.036	0.036
10^{-6}	41.7	0.870	0.05	0.036	0.029	10^{-7}	2×10^{5}	0.065	0.066	0.065	0.04
10^{-4}	3000	0.66	0.169	0.179	0.110	10^{-3}	8×10^{9}	0.225	0.2	0.12	0.15

Fig 10. shows the computation times of different techniques vs K. As expected, the solutions based on pseudo-inverse and Tikhonov regularization without the random projection have constant and rather large computation times compared to the solutions using the random projection at the smaller values of K. The computation times for the three methods with the projection are rather close, however the AMF time reaches the constant no-projection times earlier.

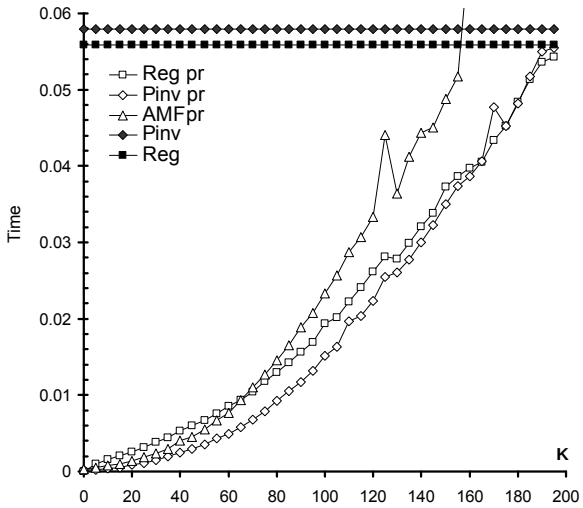

Fig 10. The computation times of studied techniques vs K.
pr denote the times with the random projection.

Conclusion

We conducted an experimental study of techniques for solving discrete ill-posed problems. The test discretized problems were those of Carasso (an inverse heat equation) and Baart (an artificial Fredholm integral equation of the first kind). We used the pseudo-inverse solution and the Tikhonov regularization with the selection of the regularization parameter λ by the discrepancy principle, and the same methods, but using preliminary projection with a random Gaussian matrix ($K \times N$).

We provided the bias-variance decomposition of solution errors for different solution techniques. Investigation of the behavior of the bias-variance terms vs the number K of the projection matrix rows showed that the bias term decreases with increasing K and the variance term increases with increasing K, so that the total error has a minimum at some K. With increasing noise level, position of the error minimum shifts to the lower values of K and the error value at the minimum increases.

With the proper choice of K, the accuracy of the pseudo-inverse with projection is at the level of or exceeds the accuracy of the Tikhonov regularization without projection for the studied test problems. In most cases, projecting reduces the solution error for Tikhonov regularization, especially significantly for the cases with large solution error before the projection. The pseudo-inverse solution error without projecting is several orders of magnitude higher than the error of the other methods, so that the pseudo-inverse solution without projection is not suitable for the considered discrete ill-posed problems.

Thus, the study and application of techniques for solving discrete ill-posed problems based on pseudo-inverse with projection is a promising direction due to their stability, manifested in the smooth change of the signal recovery error with increasing noise, and also because of the lower computational costs. The latter is due to the reduction of the computational costs of the singular value decomposition for the resulting ($N\times K$) matrix after projection, when K constitutes a small fraction of N, compared to the complexity of singular value decomposition of the original ($N\times N$) matrix. Particularly large gains are achieved at the higher noise levels because for them the optimal K is small.

Directions of further research include techniques for a computationally efficient choice of the projection matrix dimensionality K at which the solution error is close to the minimum for real situations, i.e. where the exact solution is unknown.

Bibliography

[Achlioptas, 2003] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. J. Comput. System Sci., 66(4): 671-687, 2003.

[Ailon & Chazelle, 2006] N. Ailon & B. Chazelle Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In STOC '06: Proc. 38th Ann. ACM Symp. Theory of Computing, pp. 557-563, 2006.

[Ailon & Liberty, 2008] N. Ailon & E. Liberty Fast dimension reduction using Rademacher series on dual BCH codes. In STOC '08: Proc. 40th Ann. ACM Symp. Theory of Computing, 2008

[Baart, 1982] M.L. Baart. The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned linear least-squares problems. IMA J.Numer.Anal., 2: 214-247, 1982

[Boutsidis, Drineas, & Mahoney, 2009] C. Boutsidis, P. Drineas, & M.W. Mahoney. An improved approximation algorithm for the column subset selection problem. In STOC '09: Proc. 41st Ann. ACM Symp. Theory of Computing, 2009.

[Boutsidis, Mahoney, & Drineas, 2008] C. Boutsidis, M. W. Mahoney, & P. Drineas. Unsupervised feature selection for principal components analysis. In Proc. ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining (KDD), 2008.

[Bulakh, 2006] E.G. Bulakh. A new approximation approach to solving inverse problems of gravimetry in a class of three-dimensional contact surfaces. Dokl. NASU, 1:108–112, 2006 (in Russian).

[Carasso, 1982] A.S. Carasso Determining surface temperatures from interior observations. SIAM J.Appl.Math, 42:558-574, 1982.

[Demmel, 1997] J.W. Demmel. Applied Numerical Linear Algebra. SIAM, Philadelphia. 419p., 1997.

[Deshpande et al., 2006] A. Deshpande, L. Rademacher, S. Vempala, & G. Wang. Matrix approximation and projective clustering via volume sampling. In Proc. 17th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA), pp. 1117-1126, 2006.

[Deshpande & Vempala, 2006] A. Deshpande & S. Vempala Adaptive sampling and fast low-rank matrix approximation. In Approximation, randomization and combinatorial optimization, vol. 4110 of LNCS, Springer, Berlin, pp. 292-303, 2006.

[Drineas, Kannan, & Mahoney, 2006] P. Drineas, R.Kannan, & M. W. Mahoney (Fast Monte Carlo algorithms for matrices. II. Computing a low-rank approximation to a matrix. SIAM J. Comput., 36:158-183 (electronic), 2006.

[Drineas, Kannan, & Mahoney, 2006a] P. Drineas, R.Kannan, & M. W. Mahoney. Fast Monte Carlo algorithms for matrices. III. Computing a compressed approximate matrix decomposition. SIAM J. Comput., 36:184-206, 2006

[Drineas, Mahoney, &. Muthukrishnan, 2006] P. Drineas, M. W. Mahoney, & S. Muthukrishnan. Sampling algorithms for ℓ2 regression and applications. In Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1127–1136, 2006.

[Drineas, Mahoney, & Muthukrishnan, 2008] P.Drineas, M.W. Mahoney, & S. Muthukrishnan. Relative-error CUR matrix decompositions. SIAM J. Matrix Anal. Appl., 30:844-881, 2008.

[Drineas et al., 2007] P. Drineas, M.W. Mahoney, S. Muthukrishnan, & T. Sarlos. Faster least squares approximation. Tech. Rep. 0710.1435, 2007.

[Engl, Hanke, & Neubaer, 2000] H.W. Engl, M. Hanke, & A. Neubaer. Regularization of inverse problems. Kluwer Academic Publishers, Dordrecht, 321 p., 2000.

[Goldenshluger & Pereverzev, 2000] A. Goldenshluger & S.V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. Probab. Theory Relat. Fields, 118:169–186, 2000.

[Halko, Tropp & Martinsson, 2009] N. Halko, J.A. Tropp, P.G. Martinsson Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions. In: ACM Report 2009-05, Caltech, 2009.

[Hansen, 1998] P.C. Hansen. Rank-deficient and discrete ill-posed problems. Numerical Aspects of Linear Inversion. SIAM, Philadelphia, 247 p., 1998.

[Hansen & O'Leary, 1993] P.C. Hansen & D.P. O'Leary, The use of L-curve in the regularization of discrete ill-posed problems, SIAM J.Sci.Comput. 14:1487-1503, 1993.

[Har-Peled, 2006] S. Har-Peled. Matrix approximation in linear time. Manuscript. Available at http://valis.cs.uiuc.edu/~sariel/research/papers/05/lrank/, 2006.

[Johnson & Lindenstrauss, 1984] W.B. Johnson & J. Lindenstrauss Extensions of Lipschitz mappings into a Hilbert space // Contemporary Mathematics, 26:189-206,1984.

[Khmelevskii & Bondarenko, 1999] V.K. Khmelevskii, V.M. Bondarenko. Electrical exploration prospecting. Nedra, 438 p., 1999 (in Russian).

[Li, Hastie, & Church, 2006] P. Li, T.J. Hastie, & K.W. Church Very sparse random projections. 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA: ACM Press, pp. 287-296, 2006.

[Liberty, Ailon, & Singer, 2008] E. Liberty, N. Ailon, & A. Singer. Dense fast random projections and lean Walsh transforms. In APPROX and RANDOM 2008, A. Goel et al., ed., in LNCS, Springer, Berlin, 5171:512-522, 2008.

[Liberty et al., 2007] E. Liberty, F.F. Woolfe, P.G. Martinsson, V. Rokhlin, & M. Tygert. Randomized algorithms for the low-rank approximation of matrices. Proc. Natl. Acad. Sci. USA, 104: 20167-20172, 2007.

[Misuno, Rachkovskij, & Slipchenko, 2005] I.S. Misuno, D.A. Rachkovskij, & S.V. Slipchenko. Vector and distributed representations reflecting semantic relatedness of words. Mathematical Machines and Systems, Issue 3:50–66, 2005.

[Morozov, 1984] V.A.Morozov, Methods for solving incorrectly posed problems, Springer Verlag, New York, 1984.

[Papadimitriou et al., 2000] C.H. Papadimitriou, P. Raghavan, H. Tamaki, & S. Vempala. Latent semantic indexing: A probabilistic analysis, J. Comput. System Sci., 61:217-235, 2000.

[Revunova & Rachkovskij, 2009] E.G. Revunova & D.A. Rachkovskij. Increasing the accuracy of solving the inverse problem using random projections. International Conference "Knowledge-Dialogue-Solution" (KDS-2), International Book Series "INFORMATION SCIENCE & COMPUTING", Sofia, Bulgaria: Institute FOI ITHEA, 15, pp. 93-98, 2009 (in Russian).

[Rokhlin, Szlam, & Tygert, 2009] V. Rokhlin, A. Szlam, & M. Tygert A randomized algorithm for principal component analysis. SIAM J. Matrix Anal. Appl., 31(3):1100-1124, 2009.

[Rokhlin & Tygert, 2008] V. Rokhlin & M. Tygert A fast randomized algorithm for overdetermined linear least-squares regression PNAS September, 105(36):13212-13217, 2008.

[Rudelson & Vershynin, 2007] M. Rudelson & R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. J. ACM, 54(4): article 21 (electronic), 2007.

[Ruston, 1964] A. F. Ruston. Auerbach's theorem. Math. Proc. Cambridge Philos. Soc., 56:476-480, 1964.

[Sarlos, 2006] T. Sarlos. Improved approximation algorithms for large matrices via random projections. In Proc. 47th Ann. IEEE Symp. Foundations of Computer Science (FOCS), pp. 143-152, 2006. ,

[Strakhov, 2008]. V.N. Strakhov. The fundamental computational problem in gravimetry, magnetometry, geodesy, and geoinformatics, Izvestiya Physics of the Solid Earth, 44(2):142-157, 2008.

[Tikhonov & Arsenin, 1977] A.N. Tikhonov & V.Y Arsenin. Solution of ill-posed problems. V.H. Winston, Washington, DC., 1977.

[Tygert, 2009] M. Tygert. A fast algorithm for computing minimal-norm solutions to underdetermined systems of linear equations. arXiv:0905.4745, May 2009.

[Vogel, 2002] Vogel C.R. Computational methods for inverse problems. Philadelphia: SIAM. 183 p., 2002.

[Wahba, 1990] G. Wahba, Spline models for observation data, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990.

[Woolfe et al., 2008] F. Woolfe, E. Liberty, V. Rokhlin, & M. Tygert. A fast randomized algorithm for the approximation of matrices. Appl. Comp. Harmon. Anal., 25: 335-366, 2008.

[Zabulonov, Korostil, & Revunova, 2006] Yu.L. Zabulonov, Yu.M. Korostil, E.G. Revunova Optimization of inverse problem solution to obtain the distribution density function for surface contaminations. Modeling and information technologies, 39:77-83, 2006 (in Russian).

Authors' Information

Elena G. Revunova – PhD, Research Scientist, International Research and Training Center for Information Technologies and Systems, NAS and MES of Ukraine; Pr. Acad. Glushkova, 40, Kiev, 03680, Ukraine; e-mail: egrevunova (at) gmail.com

Major Fields of Scientific Research: Signal Processing

Dmitri A. Rachkovskij – PhD, DSc, Leading Research Scientist, International Research and Training Center for Information Technologies and Systems, NAS and MES of Ukraine; Pr. Acad. Glushkova, 40, Kiev, 03680, Ukraine; e-mail: dar (at) infrm.kiev.ua

Major Fields of Scientific Research: Computational and Artificial Intelligence