Using randomization to compute the least squares solution of overdetermined linear systems by preconditioned iterations

J. Mas[*], J. Cerdán[*], J. Marín[*], and M.Tůma[†]

March 31, 2017

Abstract

We consider preconditioned iterative methods for solving large and sparse linear least squares problems. Some structures of the system matrix can seriously spoil the efficiency of standard preconditioners based on incomplete factorizations. For example dense but low rank matrices scattered around the (implicitly used) system A^TA that can appear as a result of dense rows in the system matrix $A \in \mathbb{R}^{m\times n}$ ($m \geq n$). We try to minimize the effect of such structures on the preconditioner by their shrinking based on randomization. The numerical experiments demonstrate that in this way we are able to get efficient preconditioners as well without treating a part of the matrix as an update applied via the Woodbury formula.

1 Introduction

The least squares solution of a large and sparse overdetermined inconsistent linear system

$$Ax = b, \tag{1}$$

can be computed using iterative methods. One of the most used iterative methods for LS problems is CGLS [3]. CGLS is equivalent to applying the Conjugate Gradient method (CG) to the normal equations $A^TAx = A^Tb$. Preconditioners are used to improve the convergence of the iterative method. A natural choice is to compute an incomplete Cholesky factorization of the matrix A^TA, see [5] for instance. However A^TA tends to be much denser than A —observe if A has one row whose entries are all different from 0, then A^TA is structurally full— and it is difficult to compute good and sparse preconditioners for dense matrices.

We will propose some techniques to solve the LS problems when A contains a substruture that makes impossible to compute an efficient preconditioner. To do this we propose to compute an approximate factorization of A^TA. The solution strategies are based on the problem reduction by randomization.

Recent development of the LS solvers came with a strongly innovative idea – using randomization techniques, see the seminal paper on finding structure with randomness by Halko, Martinsson and Tropp [6]. An important step using this idea in development of approximation techniques to solve the LS problem where columns of A may be close to linearly dependent can be found in [7], where the randomized matrix approximation is used to precondition the conjugate gradient method for solving the LS problems. The authors use randomized preconditioner based on randomized samplings that approximately decrease the row dimension of A. Such an approach can be efficient for A with $m \gg n$ as emphasized also in the paper [2], see also the Thesis [1], that can be considered as a continuation of the work by Rokhlin and Tygert. In [2] the authors emphasize the concept of coherence related to the quality of randomized approximation.

2 Randomization

To fix notation let $A \in \mathbb{R}^{m\times n}$, where $m \gg n$, and consider a splitting of the matrix A^TA.

$$A^TA = \sum_{i\in\mathcal{J}}^{m-k} A_i^TA_i + \sum_{i\notin\mathcal{J}} A_i^TA_i \equiv A_1^TA_1 + B^TB. \tag{2}$$

[*]Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera, 14, 46022–València(SPAIN). Email: jmasm@imm.upv.es, jmarinma@upv.es, jcerdan@imm.upv.es

[†]Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University and Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic, Email: mirektuma@karlin.mff.cuni.cz

where $\mathcal{J} \subset \{1, 2, \ldots, m\}$, and A_i is the i-th row of A and the matrix $B \in R^{m_B \times n}$, $m_B \leq n$ contains the dense rows and/or the rows that prohibit the efficient factorization of $A^T A$.

To get a cheaper algorithm we propose to approximate the range space of B^T and shrink B in this way. A set of randomization techniques that could be used for this task is described and analyzed, for example, by Halko, Martinsson and Tropp in [6]. Specifying a Gaussian random matrix Ω from $\mathbb{R}^{m_B \times k+p}$ and multiplying B^T from the right we get $Y = B^T\Omega$. If we denote orthogonalized columns of Y by Q, the corresponding projector QQ^T projects A with high probability to the space of the left singular vectors of A of its largest singular values. We will emphasize this property of the randomized choice in the following result from [6] in a significantly simplified form. Note that this theoretical guarantee is given in the form of expectation of the approximation error.

Theorem 1. *Assume that $B \in R^{m_B \times n}$ has the singular value decomposition $U_B \Sigma_B V_B^T$. Select the target row count $k, 2 \leq k \leq m_B$ and a number of samples $l, k \leq l \leq m_B$. Let Ω be a Gaussian test matrix $\Omega \in R^{n \times l}$ and Q be the orthonormal basis of the range of $B^T\Omega$ Then the expectation $\mathcal{E}||B^T - QQ^T B^T||$ with respect to Ω is at most of the size $C(k, l, m_B)\sigma_{k+1}$ where $C(k, l, m_B)$ is a small polynomial factor depending on k, l and m_B.*

Of course, σ_{k+1} is zero for any $k > m_B$ but it could happen that the numeric rank k is much smaller than m_B. Note that here we are not interested in the actual size of related constants. Instead, we would like to evaluate usefullness of this approximation in our experiments. The approximate system of normal equations obtained after the transformation can be written then as then

$$\tilde{C} = A_1^T A_1 + (Q^T B)^T Q^T Q (Q^T B) \equiv C_1 + (Q^T B)^T Q^T Q (Q^T B). \tag{3}$$

The rows of Q^T are typically dense even when the original rows of B have not been fully dense. If we accept this transformation as the only one before computing the incomplete factorization of C then we have two possible solution strategies, in general. First, when dealing with direct methods, the approximate solution $\tilde{x}$ of the LS problem can be computed using the Woodbury formula as

$$\tilde{x} \approx \tilde{C}^{-1} b = C_1^{-1} - C_1^{-1}(QQ^T B)^T (I + QQ^T B C_1^{-1}(QQ^T B)^T)^{-1} C_1^{-1} b. \tag{4}$$

We will analyze these techniques and apply them to some large problems.

Acknowledgements

This research was partially supported by Spanish Grants MTM2014-58159-P and MTM2015-68805-REDT.

References

[1] H. Avron, *Efficient and robust hybrid Direct-Iterative multipurpose linear solvers*, Ph. D. Thesis, Raymond and Beverly Sackler Faculty of Exact Sciences, School of Computer Science, Tel-Aviv University, 2010.

[2] H. Avron, P. Maymounkov, and S. Toledo, *Blendenpik: supercharging Lapack's least-squares solver*, SIAM J. Sci. Comput., 32, 3, 1217–1236 (2010), DOI: 10.1137/090767911,

[3] Å. Björck, *Numerical methods for Least Squares Problems*. SIAM, Philadelphia (1996)

[4] Å. Björck, *A general updating algorithm for constrained linear least squares problems*, SISSC, 5, 2, 394–402, 1984,

[5] R. Bru, J. Marín, J. Mas and M. Tůma, *Preconditioned Iterative Methods for Solving Linear Least Squares Problems* SIAM J. Sci. Comput., 36(4), A2002–A2022, 2014.
https://doi.org/10.1137/130931588

[6] N. Halko, P. G. Martinsson and J. A. Tropp, *Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions* SIAM Rev., 53(2), 217–288, 2011.

[7] V. Rokhlin and M. Tygert *A fast randomized algorithm for overdetermined linear least-squares regression* Proc. Natl. Acad. Sci. USA, 105 (36),13212–13217, 2008, DOI: 10.1073/pnas.0804869105

[8] J. A. Scott and M. Tuma, M., *Preconditioning of linear least squares by RIF for implicitly held normal equations*, SISC, 38, 6, C603–C623, 2016.