DESCENT SPECTRAL VERSIONS OF THE TRADITIONAL CONJUGATE GRADIENT ALGORITHMS WITH APPLICATION TO NONNEGATIVE MATRIX FACTORIZATION

Fatemeh DARGAHI[1], Saman BABAIE–KAFAKI[2],
Zohre AMINIFARD[3]

Abstract. *Despite computational superiorities, some traditional conjugate gradient algorithms such as Polak–Ribiére–Polyak and Hestenes–Stiefel methods generally fail to guarantee the descent condition. Here, in a matrix viewpoint, spectral versions of such methods are developed which fulfill the descent condition. The convergence of the given spectral algorithms is argued briefly. Afterwards, we propose an improved version of the nonnegative matrix factorization problem by adding penalty terms to the model, for controlling the condition number of one of the factorization elements. Finally, the computational merits of the method are examined using a set of CUTEr test problems as well as some random nonnegative matrix factorization models. The results typically agree with our analytical spectrum.*

Keywords: Unconstrained optimization, conjugate gradient method, spectral method, rank-one update, nonnegative matrix factorization.

DOI [10.56082/annalsarsciinfo.2024.1.35](https://doi.org/10.56082/annalsarsciinfo.2024.1.35)

1. Introduction

Scholar studies show that the introduction of conjugate gradient (CG) methods made a revolution in the field of numerical optimization. Requiring low memory and having simple iterations besides acceptable convergence, the methods have been extensively utilized in practical disciplines such as signal processing, machine learning, and neural networks training, which often appear in large-scale models.

For solving the minimization problem $\min_{x\in\mathbb{R}^n} f(x)$ with a smooth real-valued function f, here we focus on the CG methods which their search directions can be formulated by

[1] PhD student, Department of Mathematics, "Semnan University", Semnan, Iran, (fatemehdargahi@semnan.ac.ir).
[2] Prof., Department of Mathematics, "Semnan University", Semnan, Iran, (sbk@semnan.ac.ir).
[3] Senior Researcher, Department of Mathematics, "Semnan University", Semnan, Iran (aminisor@semnan.ac.ir).

$$d_0 = -g_0,\ d_{k+1} = -P_{k+1}g_{k+1},\ \text{for all } k \geq 0, \tag{1.1}$$

where $P_{k+1} \in \mathbb{R}^n$ is a rank-one update of the identity matrix as follows:

$$P_{k+1} = I - p_k q_k^T, \tag{1.2}$$

where $p_k, q_k \in \mathbb{R}^n$ are non-zero vector parameters. Notice that d_{k+1} in (1.1) can be effectively determined by performing one inner product and storing three n-dimensional vectors. Now, for the iterative method

$$x_{k+1} = x_k + s_k,\ for\ all\ k \geq 0,$$

with some $x_0 \in \mathbb{R}^n$, where $s_k = \alpha_k d_k$ with the step length $\alpha_k > 0$ fixed by a line search [30] along the direction d_k given by (1.1), the choice

$$(p_k, q_k) = \left(\frac{d_k}{\sqrt{d_k^T y_k}}, \frac{y_k}{\sqrt{d_k^T y_k}} \right), \tag{1.3}$$

with $y_k = g_{k+1} - g_k$ and $g_k = \nabla f(x_k)$, leads to the CG method proposed by Hestenes and Stiefel [17] (HS). Also, by setting

$$(p_k, q_k) = \left(\frac{d_k}{\|g_k\|}, \frac{y_k}{\|g_k\|} \right), \tag{1.4}$$

with $\|\cdot\|$ standing for the Euclidean norm, we get the method given by Polak and Ribière [23], and Polyak [24] (PRP).

From this point forward, we enforce the popular strong Wolfe conditions in the line search, i.e.

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{1.5}$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\varrho g_k^T d_k, \tag{1.6}$$

Where $0 < \delta < \varrho < 1$, ensuring $p_k^T d_k > 0$ for the choices (1.3) and (1.4) which is required in our future analysis.

Despite efficiency in the computational viewpoint among the traditional CG techniques [3], PRP and HS iterations may generate uphill directions. To combat this defect, scholars made extensive efforts to open up modification ways to get the descent versions of the methods. In this direction, a fundamental step has been taken by developing three-term CG algorithms principally by Zhang, Zhou and Li [37, 38]. In another class of modifications, researchers picked up on the thread that introduced by Dai and Liao [10] to gain descent extensions of the methods; see for example [5, 7, 16]. Spectral extensions of the PRP and HS methods also

have attracted significant attention to get the descent property; as examples take a look at the analyses carried out by Andrei [4], Dong et al. [12], Wan, Yang and Wang [33], Wan et al. [32], Faramarzi and Amini [14], Jian et al. [18], Zhang and Dan [39], Du and Liu [13], Wang et al. [34], Yu, Guan and Chen [36], and Sun et al. [31].

To make progress in the class of spectral HS/PRP algorithms, here we bring about change in the classical analyses of the literature in the sense of utilizing matrix aspects, as will be discussed in the next section (Section 2). We present our modified model for the nonnegative matrix factorization problem in Section 3. We share our numerical experiments on a category of benchmark test problems as well as some randomly generated cases of the nonnegative matrix factorization problem in Section 4. Ultimately, in Section 5 we present the concluding notes.

2. On a class of spectral descent conjugate gradient algorithms

Spectral CG algorithms have been principally introduced by Birgin and Martínez [8], reporting improved effects on the performance of the methods. Motivated by the results of [8], here we use a scalar scaling of the identity matrix in (1.2) as

$$\boldsymbol{Q_{k+1} = \mu_k I - p_k q_k^T}, \tag{2.1}$$

where the positive scalar $\boldsymbol{\mu_k}$ is called the spectral parameter, and then, we define the spectral CG directions by

$$\boldsymbol{d_0 = -g_0},\ \boldsymbol{d_{k+1} = -Q_{k+1} g_{k+1}},\ \textbf{for all } \boldsymbol{k \geq 0}. \tag{2.2}$$

Now, to avoid generating uphill search directions as an important theoretically troubling issue for several traditional CG algorithms with the search direction matrix format (1.2), we compute $\boldsymbol{\mu_k}$ in (2.1) to gain the descent property. Since $\boldsymbol{d_0^T g_0 = -\|g_0\|^2 < 0}$, and, for all $\boldsymbol{k \geq 0}$,

$$\boldsymbol{d_{k+1}^T g_{k+1} = -g_{k+1}^T Q_{k+1} g_{k+1} = -g_{k+1}^T \frac{Q_{k+1}+Q_{k+1}^T}{2} g_{k+1}}, \tag{2.3}$$

one possibility is to make the symmetric matrix

$$\boldsymbol{\mathcal{H}_{k+1} = \frac{Q_{k+1}+Q_{k+1}^T}{2} = \mu_k I - \frac{1}{2} p_k q_k^T - \frac{1}{2} q_k p_k^T},$$

to be positive definite. So, following the instructions of [6], what it takes is to determine the eigenvalues of $\mathcal{H}_{k+1}$.

Since there exists a set of $n-2$ mutually orthogonal vectors being also orthogonal to p_k and q_k, the spectral parameter μ_k is an eigenvalue of $\mathcal{H}_{k+1}$ with the multiplicity $n-2$. We pursue our analysis with the aim of finding the two other eigenvalues of $\mathcal{H}_{k+1}$ denoted by τ_k^+ and τ_k^-. In this regard, because the trace of $\mathcal{H}_{k+1}$ and the sum of its eigenvalues are equal, we get

$$\tau_k^- + \tau_k^+ = 2\mu_k - p_k^T q_k. \tag{2.4}$$

On the other hand, since $tr(\mathcal{H}_{k+1}^T \mathcal{H}_{k+1}) = \|\mathcal{H}_{k+1}\|_F^2$ with $\|\cdot\|_F$ standing for the Frobenius norm, being also equal to sum of the squared eigenvalues of $\mathcal{H}_{k+1}$, we can write

$$\tau_k^{-2} + \tau_k^{+2} = 2\mu_k^2 - 2(p_k^T q_k)\mu_k + \frac{1}{2}(p_k^T q_k)^2 + \frac{1}{2}\|p_k\|^2\|q_k\|^2. \tag{2.5}$$

Now, from (2.4) and (2.5), we get

$$\tau_k^- \tau_k^+ = \mu_k^2 - (p_k^T q_k)\mu_k + \frac{1}{4}(p_k^T q_k)^2 - \frac{1}{4}\|p_k\|^2\|q_k\|^2. \tag{2.6}$$

Thus, from (2.4) and (2.6), we should solve the following quadratic equation:

$$\tau_k^2 - (2\mu_k - p_k^T q_k)\tau_k + \mu_k^2 - (p_k^T q_k)\mu_k + \frac{1}{4}(p_k^T q_k)^2 - \frac{1}{4}\|p_k\|^2\|q_k\|^2 = 0,$$

to determine τ_k^+ and τ_k^-. Indeed, after a series of classic algebraic manipulations, we obtain

$$\tau_k^{\pm} = \frac{1}{2}(2\mu_k - p_k^T q_k \pm \|p_k\|\|q_k\|).$$

As we know, imposing positiveness on the smallest eigenvalue of $\mathcal{H}_{k+1}$ makes it positive definite. From this fact and since we can see that $\tau_k^+ \geq \mu_k \geq \tau_k^-$, next we plan to find μ_k in such a way that $\tau_k^- > 0$. So, we should have

$$\mu_k > \frac{1}{2}(p_k^T q_k + \|p_k\|\|q_k\|).$$

Thus, as a result of the above analysis, the following class of two-parameter choices for μ_k is given:

$$\mu_k = a\left(p_k^T q_k\right) + b\|p_k\|\|q_k\|, \tag{2.7}$$

Where $a, b > \frac{1}{2}$, guaranteeing the descent condition for the spectral CG methods with the directions of the general structure (2.2).

Here, abbreviations SHS and SPRP are used to represent the scaled CG algorithms respectively with the choices (1.3) and (1.4) adopted in (2.2), with μ_k defined by (2.7). For SHS, we establish the following result.

Lemma 2.1. If in (2.7) we have $a > \frac{1}{2} + \epsilon$, for some constant $\epsilon > 0$, then directions of SHS fulfill the sufficient descent condition, i.e.

$$d_k^T g_k \leq -\rho\|g_k\|^2, \textbf{ for all } k \geq 0, \tag{2.8}$$

for some $\rho > 0$.

Proof. By taking into account the relation (2.3), we get

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T \mathcal{H}_{k+1} g_{k+1} \leq -\tau_k^- \|g_{k+1}\|^2 \leq -\epsilon p_k^T q_k \|g_{k+1}\|^2.$$

Now, from (1.3) we have $p_k^T q_k = 1$ which leads to $d_{k+1}^T g_{k+1} \leq -\epsilon\|g_{k+1}\|^2$.

Next, we concisely address the convergence of SHS and SPRP using the routine measures of the literature. Henceforth, we suppose that the following assumption holds.

Assumption 2.1. For an arbitrary $x_0 \in \mathbb{R}^n$, $\Omega = \{x: f(x) \leq f(x_0)\}$ is a bounded set and in some neighborhood S of Ω, $\nabla f(x)$ is Lipschitz continuous; that is,

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \text{ for all } x, \tilde{x} \in S, \tag{2.9}$$

for some positive constant L.

Now, if f is strongly convex [30] on the neighborhood S of Ω, then, from (2.1), (2.2), and (2.9), along with Theorem 1.3.16 of [30], the sequence $\{\|d_k\|\}_{k\geq 0}$ of

SHS is bounded above. Similarly, for SPRP, if (2.8) holds and $\boldsymbol{\mu_k}$ is bounded above (ensured by setting $\boldsymbol{\mu_k \leftarrow min\{\mu_k.\ M\}}$, where M is an enough large positive constant), from Lemma 3.1 of [37] we get the boundedness of the sequence $\{\|\boldsymbol{d_k}\|\}_{\boldsymbol{k\geq 0}}$. Eventually, for both of the SHS and SPRP methods with the strong Wolfe conditions (1.5) and (1.6), Lemma 3.1 of [29] leads to the convergence in the sense of $\boldsymbol{\lim\limits_{k\to\infty} inf\|g_k\| = 0}$.

3. An improved model for the nonnegative matrix factorization problem

Among the key characteristics of the digital era, is the impact of dimensionality reduction methodologies on the analysis of large data sets. In recent years, nonnegative matrix factorization (NMF) has attracted the attention of many researchers as a simple and easily interpretable technique for extracting hidden and important features of the data [9, 20, 25, 28, 35]. Classic NMF involves (approximately) decomposing an arbitrary matrix $A \in \mathbb{R}^{m\times n}$ into two matrices $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{r\times n}$ in the sense of $A \approx WH$, under the assumption that the elements of A, W and H are nonnegative and $r \ll \min\{m, n\}$. Considering a nonnegative matrix A (i.e. $A \geq 0$ which means that the entries of A are nonnegative), modeling the NMF problem can be accomplished in the following manner:

$$\min_{W,\ H} F(W,\ H) = \frac{1}{2}\|A - WH\|_F^2, \text{ s.t. } W,H \geq 0. \tag{3.1}$$

Since being nonconvex, it is a reasonable idea to transform (3.1) into some possible convex subproblems, thereby benefiting from the advantages of the convex optimization tools. By maintaining one matrix factor constant, the other matrix factor can be calculated by solving a least-squares problem. Based on this fact, ANLS (alternating nonnegative least-squares) [22] has been traditionally regarded as an important approach, technically characterized by solving the following two least-squares models alternatively:

$$H^{k+1}=\arg\min_{H\geq 0} F(W^k, H), \tag{3.2}$$

$$W^{k+1}=\arg\min_{W\geq 0} F(W, H^{k+1}), \tag{3.3}$$

for all $k \geq 0$, starting by some initial approximations for W and H.

As known, well-conditioning is an influential issue in the field of matrix computations. So, here we focus on the condition number of the (mostly) positive definite matrix $\mathcal{H} = HH^T$ of the dimension $r \times r$ in the model (3.1) to possibly combat the collinearity between the rows of H, and as a result, to achieve the computational stability for at least one of the factorization elements. Therefore, the following improved version of the NMF model (3.1) is proposed:

$$\Im(W,\ H) = \frac{1}{2}\|A - WH\|_F^2 + \xi\kappa(\mathcal{H}), \text{ s.t. } W,H \geq 0, \tag{3.4}$$

where $\xi \geq 0$ is the penalty parameter and $\kappa(.)$ stands for the spectral condition number [26, 27].

Since generically computing $\kappa(\mathcal{H})$ in the model (3.4) is costly in the computational viewpoint as well as the CPU time, diagonal approximations of $\mathcal{H}$ are more preferable. So, we consider

$$\mathcal{H} \approx \mathcal{D} = diag(\mathcal{D}_1^*,\ \mathcal{D}_2^*,\ \cdots,\mathcal{D}_r^*),$$

where

$$\mathcal{D}_i^* = \sum_{j=1}^{n} H_{ij}^2,\ i=1, 2, \ldots, r.$$

Notably, the above estimation is derived by

$$\mathcal{D}^* = \arg\min_{D\in\mathfrak{D}^+} \|\mathcal{H} - D\|_F^2,$$

where $\mathfrak{D}^+ \in \mathbb{R}^{r\times r}$ refers to the collection of all diagonal matrices whose elements are nonnegative. In addition, we employ the following function, devised based on the relation between the geometric and arithmetic means to assess the condition number in (3.4):

$$\psi(\bar{A}) = \frac{tr(\bar{A})}{r} - \sqrt[r]{\det(\bar{A})},$$

for an arbitrary positive definite matrix $\bar{A} \in \mathbb{R}^{r\times r}$. Consequently, if $\psi(.)$ (rather than $\kappa(.)$) is applied in (3.4), it is also reasonably probable to approximate $\mathcal{H}$ in a way that its eigenvalues to be well-distributed.

As a result of our argument, and especially, in pursuit of the simplicity which is vital for large-scale cases, we propose the following modified NMF model:

$$\bar{\Im}(W,\ H) = \frac{1}{2}\|A - WH\|_F^2 + \xi\psi(\mathcal{D}), \text{s.t.}\, W, H \geq 0. \tag{3.5}$$

Additionally, the following revised versions of the least-squares models (3.2) and (3.3) should alternately be solved:

$$H^{k+1}=\arg\min_{H>0} \bar{\Im}(W^k, H),$$

$$W^{k+1}=\arg\min_{W>0} \bar{\Im}(W, H^{k+1}),$$

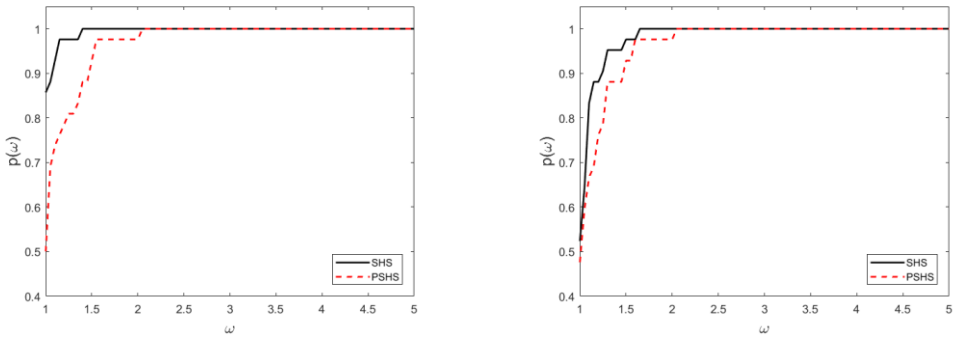
for all $k \geq 0$. The corresponding method here is called improved ANLS (IANLS) method as well. Note that the cost function (3.5) reduces to the traditional NMF cost function (3.1) when $\xi = 0$.

4. Numerical tests

To show support for our theoretical analysis, here we provide some experimental evidence across the computational spectrum. To proceed, we have selected 42 test functions of the CUTEr [15] with $n \geq 50$, as listed in Table 1. All the tests have been performed in MATLAB version 7.14.0.739 (R2012a) installed on a computer AMD FX–9800P RADEON R7, with 12 COMPUTE CORES 4C+8G 2.70 GHz of CPU and 8 GB of RAM, by the Centos 6.2 server Linux operation system. Firstly, we have compared the performance of the HS–based techniques including the SHS method and the scaled HS method of [19] (PSHS) with $\theta_k = 1$. Then, we have compared the performance of the PRP–based techniques including the SPRP method and the scaled PRP method of [33] (SPRP–WYW). We used the approximate Wolfe conditions of [16] with similar settings. The algorithms were stopped by similar conditions as given in [2] with $\|g_k\| < 10^{-5}(1 + |f_k|)$. We have scrutinized the effect of the parameters *a* and *b* in (2.7) on the performance

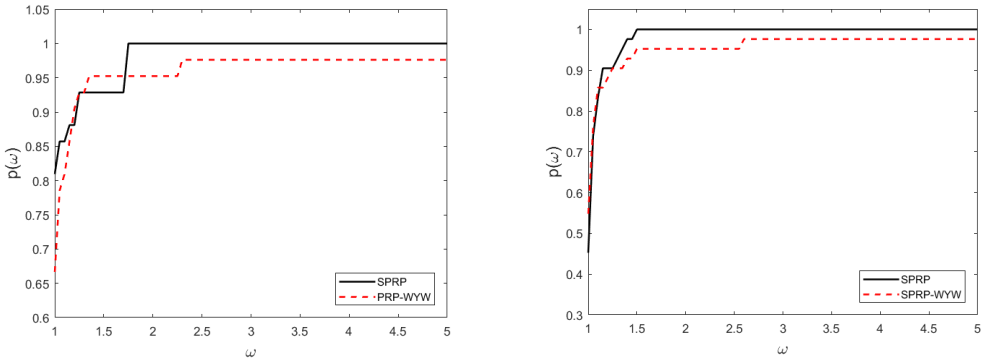
of SHS and SPRP by evaluating the outputs for different choices $a, b \in \{0 \cdot 1k\}_{k=6}^{10}$. As a result, we recognized the values $(a.\ b)$=(0.7, 0.6) for SHS and SPRP as the best.

To judge the performance of the algorithms visually, we applied the Dolan–Moré performance profile [11] following the notations of [2], on the factors of the total number of function and gradient evaluations (TNFGE) [16], and the CPU time (CPUT). Results of the performance comparisons have been depicted in Figures 1 and 2. They confirm that our matrix-based approach can be capable of delivering progress for the performance of the traditional CG algorithms such as HS and PRP.



(a) TNFGE (b) CPUT

Figure 1. Performance profile plots for SHS and PSHS



(a) TNFGE (b) CPUT

Figure 2. Performance profile plots for SPRP and SPRP–WYW

Table 1. Test functions information

Function	n	**Function**	n
ARGLINA	200	FLETCHCR	1000
BROYDN7D	5000	FMINSRF2	5625
DIXMAANA	3000	FMINSURF	5625
DQRTIC	5000	GENHUMPS	5000
EG2	1000	MANCINO	100
ENGVAL1	5000	MOREBV	5000
EXTROSNB	1000	MSQRTBLS	1024
FLETCBV2	5000	NCB20B	5000
PENALTY2	200	NONCVXU2	5000
PENALTY3	200	SINQUAD	5000
QUARTC	5000	TOINTGOR	50
SCHMVETT	5000	BQPGABIM	50
TOINTGSS	5000	BQPGASIM	50
VARDIM	200	BRATU1D	5003
VAREIGVL	50	DMN15102	66
ARGLINB	200	DMN37143	99
CURLY30	10000	DRCAV1LQ	4489
DECONVU	63	DRCAV2LQ	4489
EIGENALS	2550	DRCAV3LQ	4489
EIGENBLS	2550	FLETCBV3	5000
ERRINROS	50	FLETCHBV	5000

The final part of our experiments is devoted to evaluating the performance of SHS and SPRP for the ANLS and IANLS techniques. Setting $\xi = 1$ in the IALNS strategy, we adopted the stopping condition of [21]. Using a uniform distribution, test matrices were randomly generated for different dimensions and a similar approach was used to estimate the initial NMF elements, based on the suggestions

of [1]. Outputs have been reported in Tables 2 and 3, including the condition number (Cond) and RelErr, calculated by

$$\text{RelErr} = \frac{\|A-WH\|_F}{\|A\|_F}.$$

To summarize the outputs, it is noteworthy that condition numbers of H and RelErr show that IANLS performs better than ANLS. Therefore, it can be concluded that IANLS is capable of producing well-conditioned NMF element H that are accurate and of acceptable quality.

5. Conclusions

Probable uphill search directions may push the traditional conjugate gradient algorithms to the brink of default. To defeat such a computationally troubling issue, we developed simple spectral versions of the traditional conjugate gradient algorithms. Our study is principally based on the matrix aspects, analysing the eigenvalue features of the search direction matrix. We concisely addressed the convergence of the given methods using the common assumptions in the literature.

The focus of the next part of our research is on making possible modifications to a classic optimization model of the nonnegative matrix factorization problem. According to the available information, it is a significant problem that frequently arises across a wide range of practical contexts. Our improved model attempts to rule out the possibility of ill-conditioning in the factorization trajectory by using a classical measure function.

To examine the performance of the given methods, some computational tests have been carried out on the CUTEr functions. The outputs have been assessed based on the well-known Dolan–Moré benchmark. Furthermore, an evaluation of the quality of our improved nonnegative matrix factorization model has been conducted on several random cases. According to the results, the improved model can produce well-conditioned factorization elements with a reasonable relative error. Therefore, our theoretical assertions were supported by computational experiments.

Table 2. The outputs of SHS for NMF

Dimensions (m, n, r)	**METHOD**	**Cond** H	**RelErr**
(50, 50, 4)	ALNS	1.99E+00	6.51E-04
	IALNS	1.30E+00	6.14E-04
(100, 50, 5)	ALNS	2.05E+00	6.07E-04
	IALNS	1.20E+00	5.84E-04
(100, 100, 5)	ALNS	2.21E+00	6.45E-04
	IALNS	1.23E+00	6.44E-04
(100, 250, 5)	ALNS	2.97E+00	9.05E-04
	IALNS	1.29E+00	8.04E-04
(200, 200, 4)	ALNS	1.52E+00	7.25E-04
	IALNS	1.23E+00	7.03E-04
(200, 200, 8)	ALNS	3.26E+00	8.43E-04
	IALNS	1.32E+00	8.24E-04
(200, 300, 6)	ALNS	1.91E+00	8.75E-04
	IALNS	1.20E+00	7.50E-04

Table 3. The outputs of SPRP for NMF

Dimensions (m, n, r)	**METHOD**	**Cond** H	**RelErr**
(50, 50, 4)	ALNS	2.21E+00	7.07E-04
	IALNS	1.28E+00	6.99E-04
(100, 50, 5)	ALNS	1.52E+00	6.14E-04
	IALNS	1.26E+00	6.10E-04
(100, 100, 5)	ALNS	2.92E+00	7.71E-04
	IALNS	1.26E+00	7.50E-04
(100, 250, 5)	ALNS	4.24E+00	1.20E-03
	IALNS	1.40E+00	9.81E-04
(200, 200, 4)	ALNS	2.44E+00	7.64E-04
	IALNS	1.24E+00	7.28E-04
(200, 200, 8)	ALNS	2.19E+00	9.28E-04
	IALNS	1.29E+00	9.25E-04
(200, 300, 6)	ALNS	2.27E+00	1.09E-03
	IALNS	1.34E+00	9.60E-04

R E F E R E N C E S

[1] **[1]** M. Ahookhosh, K. Amini, and M.R. Peyghami. A nonmonotone trust-region line search method for large-scale unconstrained optimization. *APPL. MATH. MODEL.*, **36**(1):478–487, 2012.

[2] [2] Z. Aminifard and S.Babaie–Kafaki. An optimal parameter choice for the Dai–Liao family of conjugate gradient methods by avoiding a direction of the maximum magnification by the search direction matrix. *4OR*, **17**:317–330, 2019.

[3] **[3]** N. Andrei. Numerical comparison of conjugate gradient algorithms for unconstrained optimization. *STUD. INFORM. CONTROL.*, **16**(4):333–352, 2007.

[4] **[4]** N. Andrei. Scaled conjugate gradient algorithms for unconstrained optimization. *COMPUT. OPTIM. APPL.*, **38**(3):401–416, 2007.

[5] **[5]** S. Babaie–Kafaki and R. Ghanbari. A descent extension of the Polak–Ribière–Polyak conjugate gradient method. *COMPUT. MATH. APPL.*, **68**(12):2005–2011, 2014.

[6] **[6]** S. Babaie–Kafaki and R. Ghanbari. A descent family of Dai–Liao conjugate gradient methods. *OPTIM. METHODS SOFTW.*, **29**(3):583–591, 2014.

[7] **[7]** S. Babaie–Kafaki and R. Ghanbari. An optimal extension of the Polak–Ribière–Polyak conjugate gradient method. *NUMER. FUNCT. ANAL. OPTIM.*, **38**(9):1115–1124, 2017.

[8] [8] E. Birgin and J.M. Martínez. A spectral conjugate gradient method for unconstrained optimization. *APPL. MATH. OPTIM.*, **43**(2):117–128, 2001.

[9] [9] Y.C. Cho and S. Choi. Nonnegative features of spectro-temporal sounds for classification. *Pattern Recognit. Lett.*, **26**(9):1327–1336, 2005.

[10] [10] Y.H. Dai and L.Z. Liao. New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.*, **43**(1):87–101, 2001.

[11] [11] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, **91**(2, Ser. A):201–213, 2002.

[12] [12] X.L. Dong, H.W. Liu, Y.B. He, and X.M. Yang. A modified Hestenes–Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition. *J. Comput. Appl. Math.*, **281**(1):239–249, 2015.

[13] [13] X. Du and J. Liu. Global convergence of a spectral HS conjugate gradient method. *Procedia Eng.*, **15**:1487–1492, 2011.

[14] [14] P. Faramarzi and K. Amini. A modified spectral conjugate gradient method with global convergence. *J. Optim. Theory Appl.*, **182**:667–690, 2019.

[15] [15] N.I.M. Gould, D. Orban, and Ph.L. Toint. CUTEr: a constrained and

unconstrained testing environment, revisited. *ACM Trans. Math. Softw.*, **29**(4):373–394, 2003.

[16] [16] W.W. Hager and H. Zhang. Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Softw.*, **32**(1):113–137, 2006.

[17] [17] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, **49**(6):409–436, 1952.

[18] [18] J. Jian, Q. Chen, X. Jiang, Y. Zeng, and J. Yin. A new spectral conjugate gradient method for large-scale unconstrained optimization. *Optim. Methods Softw.*, **32**(3):503–515, 2017.

[19] [19] M. Khoshsimaye-Bargard and A. Ashrafi. A descent family of the spectral Hestenes–Stiefel method by considering the quasi-Newton method. *Optim. Methods Softw.*, **38**(3):495–509, 2023.

[20] [20] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, **42**(8):30–37, 2009.

[21] [21] H. Liu and X. Li. Modified subspace Barzilai–Borwein gradient method for nonnegative matrix factorization. *Comput. Optim. Appl.*, **55**(1):173–196, 2013.

[22] [22] P. Paatero and U. Tapper. Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **5**(2):111–126, 1994.

[23] [23] E. Polak and G. Ribière. Note sur la convergence de méthodes de directions conjuguées. *Rev. Fr. Inform. Rech. Opér.*, **3**(16):35–43, 1969.

[24] [24] B.T. Polyak. The conjugate gradient method in extreme problems. *USSR Comp. Math. Math. Phys.*, **9**(4):94–112, 1969.

[25] [25] F. Pompili, N. Gillis, P.A. Absil, and F. Glineur. Two algorithms for orthogonal
nonnegative matrix factorization with application to clustering. *Neurocomputing*, **141**:15–25, 2014.

[26] [26] M. Roozbeh, S. Babaie–Kafaki, and Z. Aminifard. Two penalized mixed–integer nonlinear programming approaches to tackle multicollinearity and outliers effects in linear regression models. *J. Ind. Manag. Optim.*, **17**(6):3475,
2021.

[27] [27] M. Roozbeh, S. Babaie–Kafaki, and Z. Aminifard. Improved high dimensional
regression models with matrix approximations applied to the comparative case
studies with support vector machines. *Optim. Methods Softw.*, **37**(5):1912–1929, 2022.

[28] [28] F. Shahnaz, M.W. Berry, V.P. Pauca, and R.J. Plemmons. Document clustering using nonnegative matrix factorization. *Inf. Process. Manag.*, **42**(2):373–

386, 2006.

[29] [29] K. Sugiki, Y. Narushima, and H. Yabe. Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization. *J. Optim. Theory Appl.*, **153**(3):733–757, 2012.

[30] [30] W. Sun and Y.X. Yuan. *Optimization Theory and Methods: Nonlinear Programming*. Springer, New York, 2006.

[31] [31] Z. Sun, H. Li, J. Wang, and Y. Tian. Two modified spectral conjugate gradient methods and their global convergence for unconstrained optimization. *Int. J. Comput. Math.*, **95**(10):2082–2099, 2018.

[32] [32] Z. Wan, J. Guo, J. Liu, and W. Liu. A modified spectral conjugate gradient projection method for signal recovery. *Signal Image Video Process.*, **12**:1455–1462, 2018.

[33] [33] Z. Wan, Z.L. Yang, and Y.L. Wang. New spectral PRP conjugate gradient method for unconstrained optimization. *Appl. Math. Lett.*, **24**(1):16–22, 2011.

[34] [34] L. Wang, M. Cao, F. Xing, and Y. Yang. The new spectral conjugate gradient method for large-scale unconstrained optimisation. *J. Inequal. Appl.*, **2020**(1):1–11, 2020.

[35] [35] Y. Wang, Y. Jia, C. Hu, and M. Turk. Nonnegative matrix factorization framework for face recognition. *Int. J. Pattern Recognition Artif. Intell.*, **19**(4):495–511, 2005.

[36] [36] G. Yu, L. Guan, and W. Chen. Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization. *Optim. Methods Softw.*, 23(2):275–293, 2008.

[37] [37] L. Zhang, W. Zhou, and D.H. Li. A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. *IMA J. Numer. Anal.*, **26**(4):629–640, 2006.

[38] [38] L. Zhang, W. Zhou, and D.H. Li. Some descent three-term conjugate gradient methods and their global convergence. *Optim. Methods Softw.*, **22**(4):697–711, 2007.

[39] [39] Y. Zhang and B. Dan. An efficient adaptive scaling parameter for the spectral conjugate gradient method. *Optim. Lett.*, **10**(1):119–136, 2016.