STEP SIZE DETERMINATION FOR FINDING LOW-RANK SOLUTIONS VIA NON-CONVEX BI-FACTORED MATRIX FACTORIZATION

Reinhard Panhuber and Ludger Prünte

Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR

Fraunhoferstr. 20, 53343 Wachtberg, Germany

Email: reinhard.panhuber@fhr.fraunhofer.de

ABSTRACT

In this paper we present an exact line search approach in order to find a suitable step size for the problem of recovering a low-rank matrix from linear measurements via non-convex bi-factored matrix factorization approaches as used in the bi-factored gradient descent (BFGD) algorithm. For the specific case of using the squared Frobenius norm as convex regularizer we prove that unique solutions for the step sizes exist. The computational complexity of the proposed method has the same order of magnitude than common inexact line search approaches, however, it needs only one execution of the sensing operator whereas inexact line search methods need at least two. As such our method requires less memory space and CPU time. We illustrate the functionality of the proposed method by use of simulations.

***Index Terms*—** Low-rank matrix recovery, signal processing, BFGD algorithm.

1. INTRODUCTION

Low-rank matrices appear frequently in signal processing applications e.g. in the field of remote sensing, machine learning, etc. The fundamental problem is to reconstruct an unknown low-rank matrix $\widetilde{\boldsymbol{L}} \in \mathbb{C}^{d_1 \times d_2}$ from a limited number of noisy observations of the form

$$\boldsymbol{y} = \mathcal{A}(\widetilde{\boldsymbol{L}}) + \boldsymbol{e},$$

where $\mathcal{A} : \mathbb{C}^{d_1 \times d_2} \to \mathbb{C}^n$ is a known affine transformation, $\boldsymbol{y} \in \mathbb{C}^n$ is a measurement vector, and $\boldsymbol{e} \in \mathbb{C}^n$ is additive noise. This problem can be cast as affine rank minimization (ARM) problem:

$$\begin{aligned} &\min_{\boldsymbol{L}} f(\boldsymbol{L}) \coloneqq \frac{1}{2} \left\| \boldsymbol{y} - \mathcal{A}(\boldsymbol{L}) \right\|_2^2 \\ &\text{s.t. } \operatorname{rank}(\boldsymbol{L}) \leq r, \end{aligned} \tag{1}$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm [1]. Up to now, a plethora of competing reconstruction algorithms emerged to solve (1) e. g. [1, 2, 3, 4, 5, 6, 7] to name a few. A common necessity among these approaches is the need for a singular value decomposition (SVD). However, when it comes to very big low-rank matrices an SVD becomes time consuming and furthermore memory space becomes an issue. One idea to overcome this problem is to factorize the low-rank matrix as $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{V}^{\mathrm{H}}$ with $\boldsymbol{U} \in \mathbb{C}^{d_1 \times r}$ and $\boldsymbol{V} \in \mathbb{C}^{d_2 \times r}$ which is known as Burer-Monteiro factorization. This approach reduces the number of variables to optimize, but destroys the convexity of the problem which becomes non-linear as from now on a product of unknown matrices is to be found. One particular algorithm following this approach is an iterative gradient procedure called the bi-factored gradient descent (BFGD) algorithm which comes with some convergence guarantees in case the initial solution $\boldsymbol{L}_0 = \boldsymbol{U}_0 \boldsymbol{V}_0{}^{\mathrm{H}}$ is somewhat close to $\widetilde{\boldsymbol{L}}$ [8]. It is of the form

$$\boldsymbol{U}_{t+1} = \boldsymbol{U}_t - \mu_{\boldsymbol{U}} \nabla_{\boldsymbol{U}} F\left(\boldsymbol{U}_t, \boldsymbol{V}_t\right) \tag{2}$$

$$\boldsymbol{V}_{t+1} = \boldsymbol{V}_t - \mu_{\boldsymbol{V}} \nabla_{\boldsymbol{V}} F\left(\boldsymbol{U}_t, \boldsymbol{V}_t\right), \tag{3}$$

where $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$ denote the step sizes, t indicates the iteration step, and $\nabla_{\boldsymbol{Z}}$ denotes the gradient with respect to $\boldsymbol{Z}$. For the specific case of the original objective function $f(\cdot)$ in (1) being both (restricted) strongly convex and smooth, an additional convex regularizer $g : \mathbb{C}^{r \times r} \to \mathbb{C}$ is required as

$$F\left(\boldsymbol{U}, \boldsymbol{V}\right) = f\left(\boldsymbol{U}\boldsymbol{V}^{\mathrm{H}}\right) + \lambda g\left(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} - \boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}\right), \tag{4}$$

where $\lambda \in \mathbb{R}^+$ determines the weighting of the regularization term. The additional regularizer is necessary in order to prevent $\boldsymbol{U}_t$ and $\boldsymbol{V}_t$ obtaining large condition numbers in any iteration step which is required by the BFGD algorithm to ensure convergence. It is required that:

1. g is convex and minimized at zero point i.e. $\nabla g(0) = 0$.
2. The gradient, $\nabla g\left(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} - \boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}\right) \in \mathbb{C}^{r \times r}$, is symmetric for any such pair.
3. g is μ_g-strongly convex and L_g-smooth [8].

Of particular interest is the question of how to choose the step sizes $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$. Proofs of convergence and existence of

The authors are grateful to HENSOLDT Sensors GmbH for funding the topic addressed in this paper.

suitable step sizes were shown in [8, 9, 10, 11], however, no equations based on parameters generally known in practice are given or again require an SVD. As BFGD applies a classic gradient descent scheme, classical line search strategies, as such exact and inexact line search methods, may be applied to find suitable step sizes. Exact line search methods attempt to find a step size which minimizes the cost function at current iteration for given gradients. Commonly, exact line search strategies are of limited interest since they usually bear a high computational complexity. Inexact methods, e. g. Armijo, Goldstein or Wolfe rules to name a few, try to reduce the computational burden by only validating an assumed step size, which is reduced and validated again if necessary [12].

In this paper, we propose an exact line search approach to find a suitable step size for the specific, yet very often applied case of $g(\cdot)$ being the quadratic Frobenius norm $g(\cdot) = \|\cdot\|_{\mathrm{F}}^2$. To the contrary of the described above, our exact line search approach shows the same computational complexity as Armijos, Goldsteins, or Wolfes inexact rules (and is even less demanding). To the best of the authors knowledge, such an approach was not published before.

This paper is organized as follows. The step size determination is derived in Section 2. In Section 3 a simulation is presented to illustrate the improved convergence rate. Section 4 concludes the paper.

2. STEP SIZE DETERMINATION

The step sizes $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$ are chosen such that the objective function (4) is minimized in every step subject to given gradients

$$\boldsymbol{G}_{\boldsymbol{U}} = \nabla_{\boldsymbol{U}} F(\boldsymbol{U}, \boldsymbol{V}) \tag{5}$$

$$\boldsymbol{G}_{\boldsymbol{V}} = \nabla_{\boldsymbol{V}} F(\boldsymbol{U}, \boldsymbol{V}) \tag{6}$$

as such

$$\mu_{\boldsymbol{U}} = \arg\min_{\mu} F(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V}) \tag{7}$$

$$\mu_{\boldsymbol{V}} = \arg\min_{\mu} F(\boldsymbol{U}, \boldsymbol{V} - \mu \boldsymbol{G}_{\boldsymbol{V}}). \tag{8}$$

We conduct the minimization by setting the corresponding derivatives to zero

$$\frac{\partial F(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V})}{\partial \mu} = 0 \tag{9}$$

$$\frac{\partial F(\boldsymbol{U}, \boldsymbol{V} - \mu \boldsymbol{G}_{\boldsymbol{V}})}{\partial \mu} = 0. \tag{10}$$

The derivatives in (9) and (10) are third order polynomials in μ and as such have in general three arbitrary roots which may be real positive, real negative, or complex. However, for valid step sizes, a single positive real root is needed. In the following Section 2.1 we first illustrate the derivation of (9) and (10) regarding $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$. In Section 2.2 we prove only one unique real positive root exists in the framework of the BFGD algorithm. Section 2.3 summarizes the results and states the final formulation of the step size determination.

2.1. Derivative regarding $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$

The regularized objective function (4) using the Frobenius norm for $g(\cdot)$ is in extended form

$$F(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V}) = \underbrace{\frac{1}{2}\left\|\boldsymbol{y} - \mathcal{A}\left((\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}})\boldsymbol{V}^{\mathrm{H}}\right)\right\|_2^2}_{D_{\boldsymbol{U}}}$$
$$+ \underbrace{\lambda\left\|(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}})^{\mathrm{H}}(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}) - \boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}\right\|_{\mathrm{F}}^2}_{S_{\boldsymbol{U}}} \tag{11}$$

and its gradient from (5) is given by

$$\boldsymbol{G}_{\boldsymbol{U}} = -\mathcal{A}^*(\tilde{\boldsymbol{y}})\boldsymbol{V} + 4\lambda \boldsymbol{U}\boldsymbol{X} \tag{12}$$

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - \mathcal{A}\left(\boldsymbol{U}\boldsymbol{V}^{\mathrm{H}}\right) \tag{13}$$

$$\boldsymbol{X} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} - \boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}, \tag{14}$$

where $\mathcal{A}^*(\cdot)$ is the Hermitian adjoint operator of $\mathcal{A}$. We first treat the data fidelity term $D_{\boldsymbol{U}}$ of (11), which we simplify by use of (13) to

$$D_{\boldsymbol{U}} = \frac{1}{2}\left\|\tilde{\boldsymbol{y}} + \mu \mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right)\right\|_2^2. \tag{15}$$

Its derivative is

$$\frac{\partial D_{\boldsymbol{U}}}{\partial \mu} = \mathrm{Re}\underbrace{\left\{\tilde{\boldsymbol{y}}^{\mathrm{H}}\mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right)\right\}}_{T_0} + \mu\left\|\mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right)\right\|_2^2, \tag{16}$$

where we may simplify the term T_0 further by utilizing the sum notation as

$$T_0 = \tilde{\boldsymbol{y}}^{\mathrm{H}}\mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right) = \sum_{i,j,b,r} \tilde{y}_i^* a_{ijb}^* v_{br}^* g_{\boldsymbol{U}jr} \tag{17}$$

and noticing that (12) in sum notation can be expressed as

$$g_{\boldsymbol{U}jr} = -\sum_{i,p} a_{ijp}\tilde{y}_i v_{pr} + 4\lambda \sum_m u_{jm}x_{mr}, \tag{18}$$

where $g_{\boldsymbol{U}jr} = (\boldsymbol{G}_{\boldsymbol{U}})_{jr}$ is the entry of $\boldsymbol{G}_{\boldsymbol{U}}$ at index (j, r). From (18) we can isolate the term

$$\sum_{i,p} \tilde{y}_i a_{ijp} v_{pr} = -g_{\boldsymbol{U}jr} + 4\lambda \sum_m u_{jm}x_{mr}$$

and insert it into (17) as

$$T_0 = \sum_{j,r}\left(-g_{\boldsymbol{U}jr}^* + 4\lambda \sum_m u_{jm}^* x_{mr}^*\right) g_{\boldsymbol{U}jr}$$
$$\triangleq -\|\boldsymbol{G}_{\boldsymbol{U}}\|_{\mathrm{F}}^2 + 4\lambda\left\langle \boldsymbol{X}, \boldsymbol{U}^{\mathrm{H}}\boldsymbol{G}_{\boldsymbol{U}}\right\rangle_{\mathrm{F}}, \tag{19}$$

where $\langle\cdot,\cdot\rangle_{\mathrm{F}}$ denotes the Frobenius product. Inserting (19) into (16) yields

$$\frac{\partial D_{\boldsymbol{U}}}{\partial \mu} = -\|\boldsymbol{G}_{\boldsymbol{U}}\|_{\mathrm{F}}^2 + 4\lambda\,\mathrm{Re}\left\{\left\langle \boldsymbol{X}, \boldsymbol{U}^{\mathrm{H}}\boldsymbol{G}_{\boldsymbol{U}}\right\rangle_{\mathrm{F}}\right\}$$
$$+ \mu\left\|\mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right)\right\|_2^2. \tag{20}$$

Next we treat the scale difference term $S_{\boldsymbol{U}}$ of (11), which we simplify as

$$\begin{aligned} S_{\boldsymbol{U}} &= \lambda \left\| \left(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}\right)^{\mathrm{H}} \left(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}\right) - \boldsymbol{V}^{\mathrm{H}} \boldsymbol{V} \right\|_{\mathrm{F}}^{2} \\ &= \lambda \left\| \boldsymbol{U}'^{\mathrm{H}} \boldsymbol{U}' - \boldsymbol{V}^{\mathrm{H}} \boldsymbol{V} \right\|_{\mathrm{F}}^{2} \\ &= \lambda \left\| \boldsymbol{X}' \right\|_{\mathrm{F}}^{2}, \end{aligned} \tag{21}$$

where

$$\boldsymbol{U}' = \boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}.$$

From (21) and (14) we also find

$$\boldsymbol{X}' = \boldsymbol{X} - \mu \left(\boldsymbol{U}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}} + \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{U} \right) + \mu^{2} \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}}. \tag{22}$$

At this point we introduce some abbreviations to further simplify (22) as

$$\boldsymbol{A} = \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{U} + \boldsymbol{U}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}} \tag{23}$$

$$\boldsymbol{B} = \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}}, \tag{24}$$

so

$$\boldsymbol{X}' = \boldsymbol{X} - \mu \boldsymbol{A} + \mu^{2} \boldsymbol{B}. \tag{25}$$

Due to a better handling of the Frobenius norm we again utilize the sum notation

$$\left(\boldsymbol{X}'\right)_{ij} = x'_{ij} = \sum_{s} u'^{*}_{si} u'_{sj} - \sum_{t} v^{*}_{ti} v_{tj}$$

to find the derivative as

$$\begin{aligned} \frac{\partial S_{\boldsymbol{U}}}{\partial \mu} &= \lambda \frac{\partial}{\partial \mu} \left\| \boldsymbol{X}' \right\|_{\mathrm{F}}^{2} = \lambda \frac{\partial \sum_{i,j} x'^{*}_{ij} x'_{ij}}{\partial \mu} \\ &= 2\lambda \operatorname{Re} \left\{ \sum_{i,j} \frac{\partial x'_{ij}}{\partial \mu} x'^{*}_{ij} \right\}. \end{aligned} \tag{26}$$

The required derivative is

$$\begin{aligned} \frac{\partial x'_{ij}}{\partial \mu} &= \frac{\partial \sum_{s} u'^{*}_{si} u'_{sj}}{\partial \mu} = \frac{\partial \sum_{s} \left(\boldsymbol{U}^{*} - \mu \boldsymbol{G}^{*}_{\boldsymbol{U}}\right)_{si} \left(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}\right)_{sj}}{\partial \mu} \\ &= \left(2\mu \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}} - \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{U} - \boldsymbol{U}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}} \right)_{ij} \\ &= \left(2\mu \boldsymbol{B} - \boldsymbol{A} \right)_{ij}, \end{aligned} \tag{27}$$

where at the last step we inserted the abbreviations (23) and (24). Inserting (27) into (26) yields

$$\begin{aligned} \frac{\partial S}{\partial \mu} &= 2\lambda \operatorname{Re} \left\{ \sum_{i,j} x'^{*}_{ij} \left(2\mu \boldsymbol{B} - \boldsymbol{A} \right)_{ij} \right\} \\ &= 2\lambda \operatorname{Re} \left\{ \left\langle \boldsymbol{X}', 2\mu \boldsymbol{B} - \boldsymbol{A} \right\rangle_{\mathrm{F}} \right\}. \end{aligned} \tag{28}$$

By combining (20) and (28) we have in total

$$\begin{aligned} \frac{\partial F\left(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V}\right)}{\partial \mu} &= - \left\| \boldsymbol{G}_{\boldsymbol{U}} \right\|_{\mathrm{F}}^{2} + \mu \left\| \mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}} \boldsymbol{V}^{\mathrm{H}} \right) \right\|_{2}^{2} \\ &+ \underbrace{4\lambda \operatorname{Re} \left\{ \left\langle \boldsymbol{X}, \boldsymbol{U}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}} \right\rangle_{\mathrm{F}} \right\} + 2\lambda \operatorname{Re} \left\{ \left\langle \boldsymbol{X}', 2\mu \boldsymbol{B} - \boldsymbol{A} \right\rangle_{\mathrm{F}} \right\}}_{T_1}. \end{aligned}$$

The term T_1 can be further simplified by use of (22) and carrying out some lengthy additional simplification steps which result in

$$\begin{aligned} \frac{\partial F\left(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V}\right)}{\partial \mu} &= 2\lambda \operatorname{Re} \Big\{ 2\mu^{3} \left\| \boldsymbol{B} \right\|_{\mathrm{F}}^{2} - 3\mu^{2} \left\langle \boldsymbol{A}, \boldsymbol{B} \right\rangle_{\mathrm{F}} \\ &+ \mu \left(2 \left\langle \boldsymbol{X}, \boldsymbol{B} \right\rangle_{\mathrm{F}} + \left\| \boldsymbol{A} \right\|_{\mathrm{F}}^{2} + \frac{1}{2\lambda} \left\| \mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}} \boldsymbol{V}^{\mathrm{H}} \right) \right\|_{2}^{2} \right) \\ &- \frac{1}{2\lambda} \left\| \boldsymbol{G}_{\boldsymbol{U}} \right\|_{\mathrm{F}}^{2} + \left\langle \boldsymbol{X}, \boldsymbol{C} \right\rangle_{\mathrm{F}} \Big\}, \end{aligned} \tag{29}$$

where in addition to (23) and (24) the abbreviation

$$\boldsymbol{C} = \boldsymbol{U}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{U}} - \boldsymbol{G}_{\boldsymbol{U}}^{\mathrm{H}} \boldsymbol{U} \tag{30}$$

was introduced.

The derivative regarding $\mu_{\boldsymbol{V}}$ from (10) is computed in an analogous manner to that of $\mu_{\boldsymbol{U}}$ from (9), however, some minor differences exist. The final result of this derivation is

$$\begin{aligned} \frac{\partial F\left(\boldsymbol{U}, \boldsymbol{V} - \mu \boldsymbol{G}_{\boldsymbol{V}}\right)}{\partial \mu} &= 2\lambda \operatorname{Re} \Big\{ 2\mu^{3} \left\| \boldsymbol{E} \right\|_{\mathrm{F}}^{2} - 3\mu^{2} \left\langle \boldsymbol{D}, \boldsymbol{E} \right\rangle_{\mathrm{F}} \\ &+ \mu \left(-2 \left\langle \boldsymbol{X}, \boldsymbol{E} \right\rangle_{\mathrm{F}} + \left\| \boldsymbol{D} \right\|_{\mathrm{F}}^{2} + \frac{1}{2\lambda} \left\| \mathcal{A}\left(\boldsymbol{U} \boldsymbol{G}_{\boldsymbol{V}}^{\mathrm{H}} \right) \right\|_{2}^{2} \right) \\ &- \frac{1}{2\lambda} \left\| \boldsymbol{G}_{\boldsymbol{V}} \right\|_{\mathrm{F}}^{2} + \left\langle \boldsymbol{X}, \boldsymbol{F} \right\rangle_{\mathrm{F}} \Big\}, \end{aligned} \tag{31}$$

where the gradient in (10) is given by

$$\boldsymbol{G}_{\boldsymbol{V}} = - \left(\mathcal{A}^{*} \left(\tilde{\boldsymbol{y}} \right) \right)^{\mathrm{H}} \boldsymbol{U} - 4\lambda \boldsymbol{V} \boldsymbol{X}$$

and the following abbreviations were introduced

$$\boldsymbol{D} = \boldsymbol{G}_{\boldsymbol{V}}^{\mathrm{H}} \boldsymbol{V} + \boldsymbol{V}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{V}} \tag{32}$$

$$\boldsymbol{E} = \boldsymbol{G}_{\boldsymbol{V}}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{V}} \tag{33}$$

$$\boldsymbol{F} = \boldsymbol{G}_{\boldsymbol{V}}^{\mathrm{H}} \boldsymbol{V} - \boldsymbol{V}^{\mathrm{H}} \boldsymbol{G}_{\boldsymbol{V}}. \tag{34}$$

2.2. Unique solutions for $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$

As can be seen from (29) and (31), the derivatives are third order polynomials in μ and as such have in general three arbitrary roots. However, for valid step sizes, we need a single positive real root. We prove that this is always the case in the framework of the BFGD algorithm by first showing that only one real root exists and second that this real root is positive.

Since the regularized objective function (4) is the sum of two functions $f(\cdot)$ and $g(\cdot)$ which are convex in $\boldsymbol{U}$ and $\boldsymbol{V}$, (4) is convex in $\boldsymbol{U}$ and $\boldsymbol{V}$ too. Therefore it has only one global minimum and thus only one real point where its derivative is zero. The same holds for the restricted functions in (7) and (8). The derivatives (29) and (31) thus can only have one real and two complex roots.

To prove this single real root is positive, we take a closer look at the polynomial´s coefficients. The polynomial has the general form

$$a\mu^{3} + b\mu^{2} + c\mu + d = 0.$$

By use of Descartes' rule of signs we can determine the sign of the roots, for which we only need to know the signs of the coefficients [13]. Those are shown in Tab. 1. The coefficient a is positive. The sign of coefficient b is unknown because $\boldsymbol{A}$ and $\boldsymbol{D}$ are indefinite matrices. Coefficient c is almost surely positive since $\boldsymbol{X} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} - \boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}$ is the difference of almost identical covariance matrices. Since $\boldsymbol{U}$ and $\boldsymbol{V}$ stem from a SVD as $\boldsymbol{L} = \bar{\boldsymbol{U}}\boldsymbol{\Sigma}\bar{\boldsymbol{V}}^{\mathrm{H}}$ and $\boldsymbol{U} = \bar{\boldsymbol{U}}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{V} = \bar{\boldsymbol{V}}\boldsymbol{\Sigma}^{1/2}$, $\boldsymbol{X}$ has only entries close to zero. We thus may neglect $\langle \boldsymbol{X}, \boldsymbol{B} \rangle_{\mathrm{F}}$ and $\langle \boldsymbol{X}, \boldsymbol{E} \rangle_{\mathrm{F}}$ compared to the remaining terms in (29) and (31). The same argument holds for coefficient d where we may neglect $\langle \boldsymbol{X}, \boldsymbol{C} \rangle_{\mathrm{F}}$ and $\langle \boldsymbol{X}, \boldsymbol{F} \rangle_{\mathrm{F}}$. In summary we have

$$+ |a|\, \mu^3 \pm |b|\, \mu^2 \pm |c|\, \mu - |d| = 0$$

which yields either one or three change of signs. According to Descartes' rule we have for

1 change of signs:	1 positive real root and 2 negative or complex roots.
3 changes of signs:	1 or 3 positive real roots, 2 or 0 complex roots, and 0 negative roots.

In either case we are guaranteed to have at least one positive real root Q.E.D. In case of three changes of signs we are guaranteed to only have one positive real root due to the convex objective function. By taking care that $d < 0$, we ensure to have a positive real root. The signs of b and c do not matter.

2.3. Step size determination

Since we can neglect the impact of $\boldsymbol{X}$ due to the reasons stated in Section 2.2, we may simplify (29) and (31) as

$$\frac{\partial F\left(\boldsymbol{U} - \mu \boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V}\right)}{\partial \mu} \simeq 4\lambda\mu^3 \left\|\boldsymbol{B}\right\|_{\mathrm{F}}^2 - 6\lambda\mu^2 \operatorname{Re}\left\{\langle \boldsymbol{A}, \boldsymbol{B}\rangle_{\mathrm{F}}\right\} + \mu\left(2\lambda \left\|\boldsymbol{A}\right\|_{\mathrm{F}}^2 + \left\|\mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right)\right\|_2^2\right) - \left\|\boldsymbol{G}_{\boldsymbol{U}}\right\|_{\mathrm{F}}^2 = p_{\boldsymbol{U}} \quad (35)$$

$$\frac{\partial F\left(\boldsymbol{U}, \boldsymbol{V} - \mu \boldsymbol{G}_{\boldsymbol{V}}\right)}{\partial \mu} \simeq 4\lambda\mu^3 \left\|\boldsymbol{E}\right\|_{\mathrm{F}}^2 - 6\lambda\mu^2 \operatorname{Re}\left\{\langle \boldsymbol{D}, \boldsymbol{E}\rangle_{\mathrm{F}}\right\} + \mu\left(2\lambda \left\|\boldsymbol{D}\right\|_{\mathrm{F}}^2 + \left\|\mathcal{A}\left(\boldsymbol{U}\boldsymbol{G}_{\boldsymbol{V}}^{\mathrm{H}}\right)\right\|_2^2\right) - \left\|\boldsymbol{G}_{\boldsymbol{V}}\right\|_{\mathrm{F}}^2 = p_{\boldsymbol{V}}. \quad (36)$$

The optimal step sizes (7) and (8) are thus

$$\mu_{\boldsymbol{U}} = \mathrm{real\,root\,of}\left(p_{\boldsymbol{U}}\right) \quad (37)$$

$$\mu_{\boldsymbol{V}} = \mathrm{real\,root\,of}\left(p_{\boldsymbol{V}}\right), \quad (38)$$

where $\mathrm{real\,root\,of}\,(\cdot)$ denotes an operator which yields only the pure real root of the three possible roots of (35) and (36). The computationally most expansive part in (35), exemplarily, is $\mathcal{A}\left(\boldsymbol{G}_{\boldsymbol{U}}\boldsymbol{V}^{\mathrm{H}}\right)$ in case of a random sensing operator. If we would use an inexact line search strategy like Armijos rule

$$F\left(\boldsymbol{U}_t - \mu_{\boldsymbol{U}}\boldsymbol{G}_{\boldsymbol{U}}, \boldsymbol{V}_t\right) \leq F\left(\boldsymbol{U}_t, \boldsymbol{V}_t\right) + \sigma\mu_{\boldsymbol{U}}\boldsymbol{B}, \quad (39)$$

with $\sigma \in (0, 1)$, at least two evaluations of $F(\cdot,\cdot)$ are needed and as such also two evaluations of the sensing operator $\mathcal{A}(\cdot)$. The computational complexity of (35) and (39) is $\mathcal{O}\left(d_1 d_2\,(n + r) + r^2 d_1\right)$ (neglecting $\boldsymbol{X}$ in $F(\cdot,\cdot)$ of (39) for a fair comparison), however, our proposed method needs $d_1 d_2\,(n + r) - 2r^2 d_1$ multiplications less and $d_1 d_2$ variables less to store which saves memory space and CPU time. Similar results emerge for Goldsteins or the Wolfe rules. Finally it shall be mentioned, that no SVDs in (35) and (36) are required.

3. SIMULATION RESULT

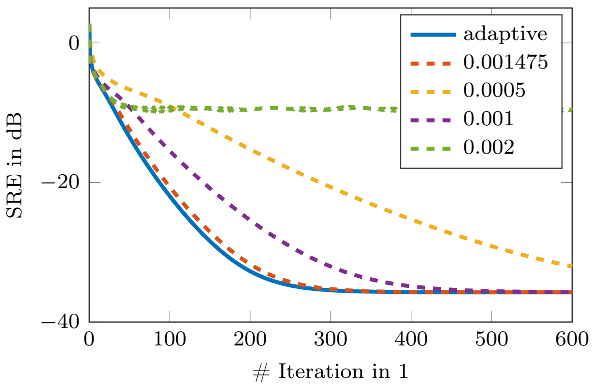
A Matlab program was designed to test the proposed step size calculation. The BFGD algorithm was used to solve (1), where the step sizes were calculated in every iteration step using (37) and (38). The dimension size was set exemplarily to $d_1 = d_2 = 70$, the rank to $r = 14$, the number of measurements to $n = d_1 d_2 / 2 = 2450$, $\lambda = 1/8$, and the entries of the affine transformation $\mathcal{A}$, $\boldsymbol{U}_{\mathrm{true}}$, and $\boldsymbol{V}_{\mathrm{true}}$ were drawn from complex standard normal distributed random samples. The columns of the matrices $\boldsymbol{U}_{\mathrm{true}}$ and $\boldsymbol{V}_{\mathrm{true}}$ were further orthogonalized and their respective ℓ_2 norms were equalized such as if $\boldsymbol{U}_{\mathrm{true}}$ and $\boldsymbol{V}_{\mathrm{true}}$ would stem from an SVD. The initial solutions $\boldsymbol{U}_0$ and $\boldsymbol{V}_0$ were obtained by perturbing $\boldsymbol{U}_{\mathrm{true}}$ and $\boldsymbol{V}_{\mathrm{true}}$ by adding substantial complex standard normal distributed noise followed by the orthogonalization and equalization procedure described before. Further, an $SNR = \left\|\mathcal{A}(\widetilde{\boldsymbol{L}})\right\|_2^2 / \left\|\boldsymbol{e}\right\|_2^2 = 40\,\mathrm{dB}$ was used. The results are shown on the top of Fig. 1 where the squared reconstruction error (SRE) metric $SRE = \left\|\boldsymbol{L} - \widetilde{\boldsymbol{L}}\right\|_{\mathrm{F}}^2 / \left\|\widetilde{\boldsymbol{L}}\right\|_{\mathrm{F}}^2$ for various step sizes and for the proposed step size determination are shown. It can be seen that our proposed approach results in best convergence rate. Furthermore, it is evident that the convergence rate is very sensitive to the chosen step size. For a step size close to the proposed one the error drops at a faster rate compared to different choices. The determined step sizes in every iteration step are shown on the bottom of Fig. 1. One may deduce from the similarity of the two graphs that it is possible to reduce the computational burden of the step size determination by using the same step size for both $\mu_{\boldsymbol{U}}$ and $\mu_{\boldsymbol{V}}$ accepting a slightly worse convergence rate. Furthermore, one may determine an optimal fixed step size from a reasonably converged step size sequence as apparent in Fig. 1 (bottom).

4. CONCLUSION

In this paper, an exact line search approach for finding low-rank solutions using the BFGD algorithm has been proposed for the specific yet very often applied case of using the squared Frobenius norm as regularizer. It is proven that only one valid real positive solution for the step size determination exists. The computational complexity of the proposed method has the same order of magnitude than common inexact line search approaches e. g. Armijo, Goldstein or Wolfe rules. However, our method is computationally less demanding as it needs only one execution of the sensing operator $\mathcal{A}$ compared to inexact strategies, which need at least two. As such our method needs less memory space and CPU time. Simulations

Coefficient	(29)	(31)	Sign
a	$4\lambda \left\| \boldsymbol{B} \right\|_{\mathrm{F}}^2$	$4\lambda \left\| \boldsymbol{E} \right\|_{\mathrm{F}}^2$	≥ 0
b	$-6\lambda \operatorname{Re}\left\{ \langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\mathrm{F}} \right\}$	$-6\lambda \operatorname{Re}\left\{ \langle \boldsymbol{D}, \boldsymbol{E} \rangle_{\mathrm{F}} \right\}$	$\gtrless 0$
c	$2\lambda \operatorname{Re}\left\{ 2 \langle \boldsymbol{X}, \boldsymbol{B} \rangle_{\mathrm{F}} + \left\| \boldsymbol{A} \right\|_{\mathrm{F}}^2 \right\} + \left\| \mathcal{A}\left(\boldsymbol{G_U} \boldsymbol{V}^{\mathrm{H}} \right) \right\|_2^2$	$2\lambda \operatorname{Re}\left\{ -2 \langle \boldsymbol{X}, \boldsymbol{E} \rangle_{\mathrm{F}} + \left\| \boldsymbol{D} \right\|_{\mathrm{F}}^2 \right\} + \left\| \mathcal{A}\left(\boldsymbol{U} \boldsymbol{G}_{\boldsymbol{V}}^{\mathrm{H}} \right) \right\|_2^2$	$\gtrsim 0$
d	$- \left\| \boldsymbol{G_U} \right\|_{\mathrm{F}}^2 + 2\lambda \operatorname{Re}\left\{ \langle \boldsymbol{X}, \boldsymbol{C} \rangle_{\mathrm{F}} \right\}$	$- \left\| \boldsymbol{G_V} \right\|_{\mathrm{F}}^2 + 2\lambda \operatorname{Re}\left\{ \langle \boldsymbol{X}, \boldsymbol{F} \rangle_{\mathrm{F}} \right\}$	$\lesssim 0$

Table 1. Signs of coefficients of (29) and (31).



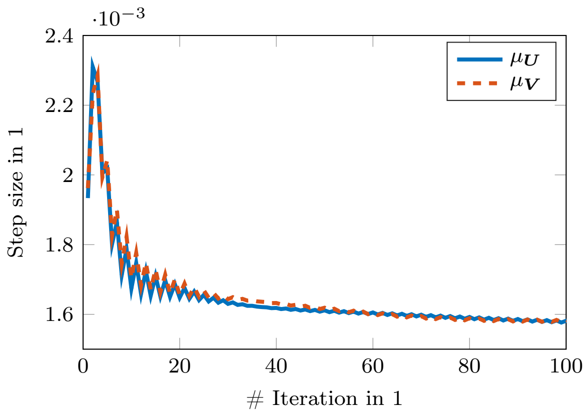


Fig. 1. Top: Convergence performance in SRE. Bottom: Step sizes determined by proposed approach.

demonstrated the functionality of the proposed method and illustrated how one can find an optimal constant step size if desired.

5. REFERENCES

[1] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 608–622, June 2016.

[2] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," 2008.

[3] M. Gavish and D. L. Donoho, "Optimal shrinkage of singular values," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2137–2152, April 2017.

[4] H. Ye, H. Li, B. Yang, F. Cao, and Y. Tang, "A novel rank approximation method for mixture noise removal of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4457–4469, July 2019.

[5] Z. Kang, C. Peng, J. Cheng, and Q. Cheng, "Logdet rank minimization with application to subspace clustering," *Computational Intelligence and Neuroscience*, vol. 2015, 2015.

[6] A. Kyrillidis and V. Cevher, "Matrix recipes for hard thresholding methods," 2012.

[7] M. Malek-Mohammadi, M. Babaie-Zadeh, and M. Skoglund, "Iterative concave rank approximation for recovering low-rank matrices," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5213–5226, Oct 2014.

[8] D. Park, A. Kyrillidis, C. Caramanis, and S. Sanghavi, "Finding low-rank solutions to matrix problems, efficiently and provably," 2016.

[9] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, "Low-rank solutions of linear matrix equations via procrustes flow," in *ICML*, 2015.

[10] X. Zhang, L. Wang, and Q. Gu, "A unified framework for nonconvex low-rank plus sparse matrix recovery," in *International Conference on Artificial Intelligence and Statistics*, 2018, pp. 1097–1107.

[11] Q. Zheng and J. Lafferty, "Convergence analysis for rectangular matrix completion using Burer-Monteiro factorization and gradient descent," 05 2016.

[12] Z. J. Shi and J. Shen, "New inexact line search method for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 127, no. 2, pp. 425–446, 2005. [Online]. Available: https://doi.org/10.1007/s10957-005-6553-6

[13] R. Descartes, *La géometrie (Discours de la méthode, third part)*, 1637, p. 373.