



HAL
open science

A proximal approach for a class of matrix optimization problems

Alessandro Benfenati, Emilie Chouzenoux, Jean-Christophe Pesquet

► **To cite this version:**

Alessandro Benfenati, Emilie Chouzenoux, Jean-Christophe Pesquet. A proximal approach for a class of matrix optimization problems. 2017. hal-01673027

HAL Id: hal-01673027

<https://hal.science/hal-01673027v1>

Preprint submitted on 28 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A PROXIMAL APPROACH FOR A CLASS OF MATRIX**
2 **OPTIMIZATION PROBLEMS***

3 ALESSANDRO BENFENATI[†], EMILIE CHOUZENOUX[‡], AND JEAN-CHRISTOPHE
4 PESQUET[‡]

5 **Abstract.** In recent years, there has been a growing interest in mathematical models leading
6 to the minimization, in a symmetric matrix space, of a Bregman divergence coupled with a regular-
7 ization term. We address problems of this type within a general framework where the regularization
8 term is split in two parts, one being a spectral function while the other is arbitrary. A Douglas-
9 Rachford approach is proposed to address such problems and a list of proximity operators is provided
10 allowing us to consider various choices for the fit-to-data functional and for the regularization term.
11 Numerical experiments show the validity of this approach for solving convex optimization problems
12 encountered in the context of sparse covariance matrix estimation. Based on our theoretical re-
13 sults, an algorithm is also proposed for noisy graphical lasso where a precision matrix has to be
14 estimated in the presence of noise. The nonconvexity of the resulting objective function is dealt with
15 a majorization-minimization approach, i.e. by building a sequence of convex surrogates and solv-
16 ing the inner optimization subproblems via the aforementioned Douglas-Rachford procedure. We
17 establish conditions for the convergence of this iterative scheme and we illustrate its good numerical
18 performance with respect to state-of-the-art approaches.

19 **Key words.** Covariance estimation, Graphical Lasso, matrix optimization, Douglas-Rachford
20 method, majorization-minimization, Bregman divergence

21 **AMS subject classifications.** 15A18, 15B48, 62J10, 65K10, 90C06, 90C25, 90C26, 90C35

22 **1. Introduction.** In recent years, various applications such as shape classifica-
23 tion models [30], gene expression [44], model selection [3, 18], computer vision [33],
24 inverse covariance estimation [31, 29, 68, 28, 62], graph estimation [48, 53, 67], social
25 network and corporate inter-relationships analysis [2], or brain network analysis [65]
26 have led to matrix variational formulations of the form:

27 (1)
$$\underset{\mathbf{C} \in \mathcal{S}_n}{\text{minimize}} \ f(\mathbf{C}) - \text{trace}(\mathbf{T}\mathbf{C}) + g(\mathbf{C}),$$

28 where \mathcal{S}_n is the set of real symmetric matrices of dimension $n \times n$, \mathbf{T} is a given
29 $n \times n$ real matrix (without loss of generality, it will be assumed to be symmetric), and
30 $f: \mathcal{S}_n \rightarrow]-\infty, +\infty]$ and $g: \mathcal{S}_n \rightarrow]-\infty, +\infty]$ are lower-semicontinuous functions
31 which are proper, in the sense that they are finite at least in one point.

32 It is worth noticing that the notion of Bregman divergence [13] gives a particular
33 insight into Problem (1). Indeed, suppose that f is a convex function differentiable
34 on the interior of its domain $\text{int}(\text{dom } f) \neq \emptyset$. Let us recall that, in \mathcal{S}_n endowed with
35 the Frobenius norm, the f -Bregman divergence between $\mathbf{C} \in \mathcal{S}_n$ and $\mathbf{Y} \in \text{int}(\text{dom } f)$
36 is

37 (2)
$$D^f(\mathbf{C}, \mathbf{Y}) = f(\mathbf{C}) - f(\mathbf{Y}) - \text{trace}(\mathbf{T}(\mathbf{C} - \mathbf{Y})),$$

*Submitted to the editors DATE.

Funding: This work was funded by the Agence Nationale de la Recherche under grant ANR-14-CE27-0001 GRAPHSSIP.

[†]Laboratoire d'Informatique Gaspard Monge, ESIEE Paris, University Paris-Est, FR (alessandro.benfenati@esiee.fr).

[‡]Center for Visual Computing, INRIA Saclay and CentraleSupélec, University Paris-Saclay, FR (emilie.chouzenoux@centralesupelec.fr, jean-christophe@pesquet.eu).

38 where $\mathbf{T} = \nabla f(\mathbf{Y})$ is the gradient of f at \mathbf{Y} . Hence, the original problem (1) is
 39 equivalently expressed as

$$40 \quad (3) \quad \underset{\mathbf{C} \in \mathcal{S}_n}{\text{minimize}} \quad g(\mathbf{C}) + D^f(\mathbf{C}, \mathbf{Y}).$$

41 Solving Problem (3) amounts to computing the proximity operator of g at \mathbf{Y} with
 42 respect to the divergence D^f [5, 7] in the space \mathcal{S}_n . In the vector case, such kind
 43 of proximity operator has been found to be useful in a number of recent works re-
 44 garding, for example, image restoration [14, 8, 9, 70], image reconstruction [71], and
 45 compressive sensing problems [66, 32].

46 In this paper, it will be assumed that f belongs to the class of *spectral functions* [11,
 47 Chapter 5, Section 2], i.e., for every permutation matrix $\Sigma \in \mathbb{R}^{n \times n}$,

$$48 \quad (4) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad f(\mathbf{C}) = \varphi(\Sigma \mathbf{d}),$$

49 where $\varphi: \mathbb{R}^n \rightarrow]-\infty, +\infty]$ is a proper lower semi-continuous convex function and \mathbf{d}
 50 is a vector of eigenvalues of \mathbf{C} .

51 Due to the nature of the problems, in many of the aforementioned applications, g is a
 52 regularization function promoting the sparsity of \mathbf{C} . We consider here a more generic
 53 class of regularization functions obtained by decomposing g as $g_0 + g_1$, where g_0 is a
 54 spectral function, i.e., for every permutation matrix $\Sigma \in \mathbb{R}^{n \times n}$,

$$55 \quad (5) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad g_0(\mathbf{C}) = \psi(\Sigma \mathbf{d}),$$

56 with $\psi: \mathbb{R}^n \rightarrow]-\infty, +\infty]$ a proper lower semi-continuous function, \mathbf{d} still denoting
 57 a vector of the eigenvalues of \mathbf{C} , while $g_1: \mathcal{S}_n \rightarrow]-\infty, +\infty]$ is a proper lower semi-
 58 continuous function which cannot be expressed under a spectral form.

59 A very popular and useful example encompassed by our framework is the graph-
 60 ical lasso (GLASSO) problem, where f is the minus log-determinant function, g_1
 61 is a component-wise ℓ_1 norm (of the matrix elements), and $g_0 \equiv 0$. Various algo-
 62 rithms have been proposed to solve Problem (1) in this context, including the popular
 63 GLASSO algorithm [31] and some of its recent variants [47]. We can also mention the
 64 dual block coordinate ascent method from [3], the SPICE algorithm [57], the gradi-
 65 ent projection method in [30], the Refitted CLIME algorithm [17], various algorithms
 66 [28, 42, 43] based on Nesterov's smooth gradient approach [50], ADMM approaches
 67 [68, 58], an inexact Newton method [62], and interior point methods [67, 40]. A re-
 68 lated model is addressed in [44, 18], with the additional assumption that the sought
 69 solution can be split as $\mathbf{C}_1 + \mathbf{C}_2$, where \mathbf{C}_1 is sparse and \mathbf{C}_2 is low-rank. Finally, let
 70 us mention the ADMM algorithm from [72], and the incremental proximal gradient
 71 approach from [54], both addressing Problem (1) when f is the squared Frobenius
 72 norm, g_0 is a nuclear norm, and g_1 is an element-wise ℓ_1 norm.

73 The main goal of this paper is to propose numerical approaches for solving Prob-
 74 lem (1). Two settings will be investigated, namely (i) $g_1 \equiv 0$, i.e. the whole cost
 75 function is a spectral one, (ii) $g_1 \neq 0$. In the former case, some general results
 76 concerning the D^f -proximity operator of g_0 are established. In the latter case, a
 77 Douglas-Rachford optimization method is proposed, which leads us to calculate the
 78 proximity operators of several spectral functions of interest. We then consider ap-
 79 plications of our results to the estimation of (possibly low-rank) covariance matrices
 80 from noisy observations of multivalued random variables. Two variational approaches
 81 are proposed for estimating the unknown covariance matrix, depending on the prior
 82 assumptions made on it. We show that the cost function arising from the first for-
 83 mulation can be minimized through our proposed Douglas-Rachford procedure under

84 mild assumptions on the involved regularization functions. The second formulation of
 85 the problem aims at preserving desirable sparsity properties of the inverse covariance
 86 (i.e., precision) matrix. We establish that the proposed objective function is a dif-
 87 ference of convex terms, and we introduce a novel majorization-minimization (MM)
 88 algorithm to optimize it.

89 The paper is organized as follows. Section 2 is devoted to the solution of the
 90 particular instance of Problem (1) corresponding to $g_1 \equiv 0$. Section 3 describes a
 91 proximal to address the problem when $g_1 \not\equiv 0$. Its implementation is discussed for
 92 a bunch of useful choices for the involved functionals. Section 4 presents two new
 93 approaches for estimating covariance matrices from noisy data. Finally, in Section 5,
 94 numerical experiments illustrate the applicability of the proposed methods, and its
 95 good performance with respect to the state-of-the-art, in two distinct scenarios.

96 *Notation:* Greek letters usually designate real numbers, bold letters designate
 97 vectors in a Euclidean space, capital bold letters indicate matrices. The i -th element
 98 of the vector \mathbf{d} is denoted by d_i . $\text{Diag}(\mathbf{d})$ denotes the diagonal matrix whose diagonal
 99 elements are the components of \mathbf{d} . \mathcal{D}_n is the cone of vectors $\mathbf{d} \in \mathbb{R}^n$ whose components
 100 are ordered by decreasing values. The symbol $\text{vect}(\mathbf{C})$ denotes the vector resulting
 101 from a column-wise ordering of the elements of matrix \mathbf{C} . The product $\mathbf{A} \otimes \mathbf{B}$ denotes
 102 the classical Kronecker product of matrices \mathbf{A} and \mathbf{B} . Let \mathcal{H} be a real Hilbert space
 103 endowed with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\|\cdot\|$, the domain of a function $f: \mathcal{H} \rightarrow$
 104 $] -\infty, +\infty]$ is $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$. f is coercive if $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$
 105 and supercoercive if $\lim_{\|x\| \rightarrow +\infty} f(x)/\|x\| = +\infty$. The Moreau subdifferential of f at
 106 $x \in \mathcal{H}$ is $\partial f(x) = \{t \in \mathcal{H} \mid (\forall y \in \mathcal{H}) f(y) \geq f(x) + \langle t, y - x \rangle\}$. $\Gamma_0(\mathcal{H})$ denotes the class of
 107 lower-semicontinuous convex functions from \mathcal{H} to $] -\infty, +\infty]$ with a nonempty domain
 108 (proper). If $f \in \Gamma_0(\mathcal{H})$ is (Gâteaux) differentiable at $x \in \mathcal{H}$, then $\partial f(x) = \{\nabla f(x)\}$
 109 where $\nabla f(x)$ is the gradient of f at x . If a function $f: \mathcal{H} \rightarrow] -\infty, +\infty]$ possesses a
 110 unique minimizer on a set $E \subset \mathcal{H}$, it will be denoted by $\underset{x \in E}{\text{argmin}} f(x)$. If there are

111 possibly several minimizers, their set will be denoted by $\underset{x \in E}{\text{Argmin}} f(x)$. Given a set E ,
 112 $\text{int}(E)$ designates the interior of E and ι_E denotes the indicator function of the set,
 113 which is equal to 0 over this set and $+\infty$ otherwise. In the remainder of the paper, the
 114 underlying Hilbert space will be \mathcal{S}_n , the set of real symmetric matrices equipped with
 115 the Frobenius norm, denoted by $\|\cdot\|_F$. The matrix spectral norm is denoted by $\|\cdot\|_S$,
 116 the ℓ_1 norm of a matrix $\mathbf{A} = (A_{i,j})_{i,j}$ is $\|\mathbf{A}\|_1 = \sum_{i,j} |A_{i,j}|$. For every $p \in [1, +\infty[$,
 117 $\mathcal{R}_p(\cdot)$ denotes the Schatten p -norm, the nuclear norm being obtained when $p = 1$.
 118 \mathcal{O}_n denotes the set of orthogonal matrices of dimension n with real elements; \mathcal{S}_n^+ and
 119 \mathcal{S}_n^{++} denote the set of real symmetric positive semidefinite, and symmetric positive
 120 definite matrices, respectively, of dimension n . \mathbf{I}_d denotes the identity matrix whose
 121 dimension will be clear from the context. The soft thresholding operator soft_μ and
 122 the hard thresholding operator hard_μ of parameter $\mu \in [0, +\infty[$ are given by

$$123 \quad (6) \quad (\forall \xi \in \mathbb{R}) \quad \text{soft}_\mu(\xi) = \begin{cases} \xi - \mu & \text{if } \xi > \mu \\ \xi + \mu & \text{if } \xi < -\mu \\ 0 & \text{otherwise} \end{cases}, \quad \text{hard}_\mu(\xi) = \begin{cases} \xi & \text{if } |\xi| > \mu \\ 0 & \text{otherwise.} \end{cases}$$

124 **2. Spectral Approach.** In this section, we show that, in the particular case
 125 when $g_1 \equiv 0$, Problem (1) reduces to the optimization of a function defined on \mathbb{R}^n .
 126 Indeed, the problem then reads:

$$127 \quad (7) \quad \underset{\mathbf{C} \in \mathcal{S}_n}{\text{minimize}} f(\mathbf{C}) - \text{trace}(\mathbf{TC}) + g_0(\mathbf{C}),$$

128 where the spectral forms of f and g_0 allow us to take advantage of the eigendecom-
129 positions of \mathbf{C} and \mathbf{T} in order to simplify the optimization problem, as stated below.

130 **THEOREM 2.1.** *Let $\mathbf{t} \in \mathbb{R}^n$ be a vector of eigenvalues of \mathbf{T} and let $\mathbf{U}_{\mathbf{T}} \in \mathcal{O}_n$
131 be such that $\mathbf{T} = \mathbf{U}_{\mathbf{T}} \text{Diag}(\mathbf{t}) \mathbf{U}_{\mathbf{T}}^{\top}$. Let f and g_0 be functions satisfying (4) and (5),
132 respectively, where φ and ψ are lower-semicontinuous functions. Assume that $\text{dom } \varphi \cap$
133 $\text{dom } \psi \neq \emptyset$ and that the function $\mathbf{d} \mapsto \varphi(\mathbf{d}) - \mathbf{d}^{\top} \mathbf{t} + \psi(\mathbf{d})$ is coercive. Then a solution
134 to Problem (7) exists, which is given by*

$$135 \quad (8) \quad \widehat{\mathbf{C}} = \mathbf{U}_{\mathbf{T}} \text{Diag}(\widehat{\mathbf{d}}) \mathbf{U}_{\mathbf{T}}^{\top}$$

136 where $\widehat{\mathbf{d}}$ is any solution to the following problem:

$$137 \quad (9) \quad \underset{\mathbf{d} \in \mathbb{R}^n}{\text{minimize}} \quad \varphi(\mathbf{d}) - \mathbf{d}^{\top} \mathbf{t} + \psi(\mathbf{d}).$$

138 For the sake of clarity, before establishing this result, we recall two useful lemmas
139 from linear algebra.

140 **LEMMA 2.2.** [46, Chapter 9, Sec. H, p. 340] *Let $\mathbf{C} \in \mathcal{S}_n$ and let $\mathbf{d} \in \mathcal{D}_n$ be a
141 vector of ordered eigenvalues of this matrix. Let $\mathbf{T} \in \mathcal{S}_n$ and let $\mathbf{t} \in \mathcal{D}_n$ be a vector of
142 ordered eigenvalues of this matrix. The following inequality holds:*

$$143 \quad (10) \quad \text{trace}(\mathbf{C}\mathbf{T}) \leq \mathbf{d}^{\top} \mathbf{t}.$$

144 *In addition, the upper bound is reached if and only if \mathbf{T} and \mathbf{C} share the same eigen-
145 basis, i.e. there exists $\mathbf{U} \in \mathcal{O}_n$ such that $\mathbf{C} = \mathbf{U} \text{Diag}(\mathbf{d}) \mathbf{U}^{\top}$ and $\mathbf{T} = \mathbf{U} \text{Diag}(\mathbf{t}) \mathbf{U}^{\top}$.*

146 The subsequent lemma is also known as the *rearrangement inequality*:

147 **LEMMA 2.3.** [34, Section 10.2, Theorem 368] *Let $\mathbf{a} \in \mathcal{D}_n$ and $\mathbf{b} \in \mathcal{D}_n$. Then, for
148 every permutation matrix \mathbf{P} of dimension $n \times n$,*

$$149 \quad (11) \quad \mathbf{a}^{\top} \mathbf{P} \mathbf{b} \leq \mathbf{a}^{\top} \mathbf{b}.$$

150 We are now ready to prove [Theorem 2.1](#).

Proof of Theorem 2.1. Due to the assumptions made on f and g_0 , Problem (7)
can be reformulated as

$$\underset{\mathbf{d} \in \mathcal{D}_n, \mathbf{U}_{\mathbf{C}} \in \mathcal{O}_n}{\text{minimize}} \quad \varphi(\mathbf{d}) - \text{trace}(\mathbf{U}_{\mathbf{C}} \text{Diag}(\mathbf{d}) \mathbf{U}_{\mathbf{C}}^{\top} \mathbf{T}) + \psi(\mathbf{d}).$$

According to the first claim in [Lemma 2.2](#),

$$\inf_{\mathbf{d} \in \mathcal{D}_n, \mathbf{U}_{\mathbf{C}} \in \mathcal{O}_n} \varphi(\mathbf{d}) - \text{trace}(\mathbf{U}_{\mathbf{C}} \text{Diag}(\mathbf{d}) \mathbf{U}_{\mathbf{C}}^{\top} \mathbf{T}) + \psi(\mathbf{d}) \geq \inf_{\mathbf{d} \in \mathcal{D}_n} \varphi(\mathbf{d}) - \mathbf{d}^{\top} \tilde{\mathbf{t}} + \psi(\mathbf{d}),$$

151 where $\tilde{\mathbf{t}} \in \mathcal{D}_n$ is the vector of ordered eigenvalues of $\mathbf{T} = \tilde{\mathbf{U}}_{\mathbf{T}} \text{Diag}(\tilde{\mathbf{t}}) \tilde{\mathbf{U}}_{\mathbf{T}}^{\top}$ with $\tilde{\mathbf{U}}_{\mathbf{T}} \in$
152 \mathcal{O}_n . In addition, the last claim in [Lemma 2.2](#) allows us to conclude that the lower
153 bound is attained when $\mathbf{U}_{\mathbf{C}} = \tilde{\mathbf{U}}_{\mathbf{T}}$. This proves that

$$154 \quad (12) \quad \inf_{\mathbf{C} \in \mathcal{S}_n} f(\mathbf{C}) - \text{trace}(\mathbf{T}\mathbf{C}) + g_0(\mathbf{C}) = \inf_{\mathbf{d} \in \mathcal{D}_n} \varphi(\mathbf{d}) - \mathbf{d}^{\top} \tilde{\mathbf{t}} + \psi(\mathbf{d}).$$

155 Let us now show that ordering the eigenvalues is unnecessary for our purposes. Let $\mathbf{t} \in$
156 \mathbb{R}^n be a vector of non necessarily ordered eigenvalues of \mathbf{T} . Then, $\mathbf{T} = \mathbf{U}_{\mathbf{T}} \text{Diag}(\mathbf{t}) \mathbf{U}_{\mathbf{T}}^{\top}$

157 with $\mathbf{U}_T \in \mathcal{O}_n$ and there exists a permutation matrix \mathbf{Q} such that $\mathbf{t} = \mathbf{Q}\tilde{\mathbf{t}}$. For every
 158 vector $\mathbf{d} \in \mathcal{D}_n$ and for every permutation matrix \mathbf{P} of dimension $n \times n$, we have then

$$\begin{aligned}
 159 \quad (13) \quad \varphi(\mathbf{P}\mathbf{d}) - (\mathbf{P}\mathbf{d})^\top \mathbf{t} + \psi(\mathbf{P}\mathbf{d}) &= \varphi(\mathbf{P}\mathbf{d}) - (\mathbf{P}\mathbf{d})^\top \mathbf{Q}\tilde{\mathbf{t}} + \psi(\mathbf{P}\mathbf{d}) \\
 160 &= \varphi(\mathbf{d}) - (\mathbf{Q}^\top \mathbf{P}\mathbf{d})^\top \tilde{\mathbf{t}} + \psi(\mathbf{d}) \\
 161 &\geq \varphi(\mathbf{d}) - \mathbf{d}^\top \tilde{\mathbf{t}} + \psi(\mathbf{d}),
 \end{aligned}$$

163 where the last inequality is a direct consequence of [Lemma 2.3](#). In addition, the
 164 equality is obviously reached if $\mathbf{P} = \mathbf{Q}$. Since every vector in \mathbb{R}^n can be expressed as
 165 permutation of a vector in \mathcal{D}_n , we deduce that

$$166 \quad (14) \quad \inf_{\mathbf{d} \in \mathbb{R}^n} \varphi(\mathbf{d}) - \mathbf{d}^\top \mathbf{t} + \psi(\mathbf{d}) = \inf_{\mathbf{d} \in \mathcal{D}_n} \varphi(\mathbf{d}) - \mathbf{d}^\top \tilde{\mathbf{t}} + \psi(\mathbf{d}).$$

167 Altogether, (12) and (14) lead to

$$168 \quad (15) \quad \inf_{\mathbf{C} \in \mathcal{S}_n} f(\mathbf{C}) - \text{trace}(\mathbf{T}\mathbf{C}) + g_0(\mathbf{C}) = \inf_{\mathbf{d} \in \mathbb{R}^n} \varphi(\mathbf{d}) - \mathbf{d}^\top \mathbf{t} + \psi(\mathbf{d}).$$

169 Since the function $\mathbf{d} \mapsto \varphi(\mathbf{d}) - \mathbf{d}^\top \mathbf{t} + \psi(\mathbf{d})$ is proper, lower-semicontinuous, and
 170 coercive, it follows from [\[56, Theorem 1.9\]](#) that there exists $\hat{\mathbf{d}} \in \mathbb{R}^n$ such that

$$171 \quad (16) \quad \varphi(\hat{\mathbf{d}}) - \hat{\mathbf{d}}^\top \mathbf{t} + \psi(\hat{\mathbf{d}}) = \inf_{\mathbf{d} \in \mathbb{R}^n} \varphi(\mathbf{d}) - \mathbf{d}^\top \mathbf{t} + \psi(\mathbf{d}).$$

172 In addition, it is easy to check that if $\hat{\mathbf{C}}$ is given by (8) then

$$173 \quad (17) \quad f(\hat{\mathbf{C}}) - \text{trace}(\mathbf{T}\hat{\mathbf{C}}) + g_0(\hat{\mathbf{C}}) = \varphi(\hat{\mathbf{d}}) - \hat{\mathbf{d}}^\top \mathbf{t} + \psi(\hat{\mathbf{d}}),$$

174 which yields the desired result. □

175 Before deriving a main consequence of this result, we need to recall some definitions
 176 from convex analysis [\[55, Chapter 26\]](#) [\[5, Section 3.4\]](#):

177 **DEFINITION 2.4.** *Let \mathcal{H} be a finite dimensional real Hilbert space with norm $\|\cdot\|$
 178 and scalar product $\langle \cdot, \cdot \rangle$. Let $h: \mathcal{H} \rightarrow]-\infty, +\infty]$ be a proper convex function.*

- 179 • *h is essentially smooth if h is differentiable on $\text{int}(\text{dom } h) \neq \emptyset$ and*
 180 *$\lim_{n \rightarrow +\infty} \|\nabla h(x_n)\| = +\infty$ for every sequence $(x_n)_{n \in \mathbb{N}}$ of $\text{int}(\text{dom } h)$ con-*
 181 *verging to a point on the boundary of $\text{dom } h$.*
- 182 • *h is essentially strictly convex if h is strictly convex on every convex subset*
 183 *of the domain of its subdifferential.*
- 184 • *h is a Legendre function if it is both essentially smooth and essentially strictly*
 185 *convex.*
- 186 • *If h is differentiable on $\text{int}(\text{dom } h) \neq \emptyset$, the h -Bregman divergence is the*
 187 *function D^h defined on \mathcal{H}^2 as*

$$\begin{aligned}
 188 \quad (18) \quad (\forall (x, y) \in \mathcal{H}^2) \\
 189 \\
 190 \quad D^h(x, y) &= \begin{cases} h(x) - h(y) - \langle \nabla h(y), x - y \rangle & \text{if } y \in \text{int}(\text{dom } f) \\ +\infty & \text{otherwise.} \end{cases} \\
 191
 \end{aligned}$$

- 192 • *Assume that h is a lower-semicontinuous Legendre function and that ℓ is*
 193 *a lower-semicontinuous convex function such that $\text{int}(\text{dom } h) \cap \text{dom } \ell \neq \emptyset$*

194 and either ℓ is bounded from below or $h + \ell$ is supercoercive. Then, the D^h -
195 proximity operator of ℓ is

$$196 \quad (19) \quad \text{prox}_\ell^h : \text{int}(\text{dom } h) \rightarrow \text{int}(\text{dom } h) \cap \text{dom } \ell$$

$$197 \quad y \mapsto \underset{x \in \mathcal{H}}{\text{argmin}} \ell(x) + D^h(x, y).$$

199 In this definition, when $h = \|\cdot\|^2/2$, we recover the classical definition of the proximity
200 operator in [49], which is defined over \mathcal{H} , for every function $\ell \in \Gamma_0(\mathcal{H})$, and that will
201 be simply denoted by prox_ℓ .

202 We will also need the following result:

203 LEMMA 2.5. Let f be a function satisfying (4) where $\varphi: \mathbb{R}^n \rightarrow]-\infty, +\infty]$. Let
204 $\mathbf{C} \in \mathcal{S}_n$ and let $\mathbf{d} \in \mathbb{R}^n$ be a vector of eigenvalues of this matrix. The following hold:

- 205 (i) $\mathbf{C} \in \text{dom } f$ if and only if $\mathbf{d} \in \text{dom } \varphi$;
206 (ii) $\mathbf{C} \in \text{int}(\text{dom } f)$ if and only if $\mathbf{d} \in \text{int}(\text{dom } \varphi)$.

207 *Proof.* (i) obviously holds since f is a spectral function.

208 Let us now prove (ii). If $\mathbf{C} \in \text{int}(\text{dom } f)$, then $\mathbf{d} \in \text{dom } \varphi$. In addition, there exists
209 $\rho \in]0, +\infty[$ such that, for every $\mathbf{C}' \in \mathcal{S}_n$, if $\|\mathbf{C}' - \mathbf{C}\|_{\text{F}} \leq \rho$, then $\mathbf{C}' \in \text{dom } f$. Let
210 $\mathbf{U}_{\mathbf{C}} \in \mathcal{O}_n$ be such that $\mathbf{C} = \mathbf{U}_{\mathbf{C}} \text{Diag}(\mathbf{d})\mathbf{U}_{\mathbf{C}}^{\top}$ and let us choose $\mathbf{C}' = \mathbf{U}_{\mathbf{C}} \text{Diag}(\mathbf{d}')\mathbf{U}_{\mathbf{C}}^{\top}$
211 with $\mathbf{d}' \in \mathbb{R}^n$. Since \mathbf{C} and \mathbf{C}' share the same eigenbasis,

$$212 \quad (20) \quad \|\mathbf{C}' - \mathbf{C}\|_{\text{F}} = \|\mathbf{d}' - \mathbf{d}\|.$$

213 Hence, for any $\mathbf{d}' \in \mathbb{R}^n$ such that $\|\mathbf{d}' - \mathbf{d}\| \leq \rho$, $\mathbf{C}' \in \text{dom } f$, hence $\mathbf{d}' \in \text{dom } \varphi$. This
214 shows that $\mathbf{d} \in \text{int}(\text{dom } \varphi)$.

215 Conversely, let us assume that $\mathbf{d} = (d_i)_{1 \leq i \leq n} \in \text{int}(\text{dom } \varphi)$. Without loss of generality,
216 it can be assumed that $\mathbf{d} \in \mathcal{D}_n$. There thus exists $\rho \in]0, +\infty[$ such that for every
217 $\mathbf{d}' = (d'_i)_{1 \leq i \leq n} \in \mathcal{D}_n$, if

$$218 \quad (21) \quad (\forall i \in \{1, \dots, n\}) \quad |d'_i - d_i| \leq \rho,$$

219 then $\mathbf{d}' \in \text{dom } \varphi$. Furthermore, let \mathbf{C}' be any matrix in \mathcal{S}_n such that

$$220 \quad (22) \quad \|\mathbf{C}' - \mathbf{C}\|_{\text{F}} \leq \rho$$

221 and let $\mathbf{d}' = (d'_i)_{1 \leq i \leq n} \in \mathcal{D}_n$ be a vector of eigenvalues of \mathbf{C}' . It follows from Weyl's
222 inequality [46] that

$$223 \quad (23) \quad (\forall i \in \{1, \dots, n\}) \quad |d'_i - d_i| \leq \|\mathbf{C}' - \mathbf{C}\|_{\text{S}} \leq \|\mathbf{C}' - \mathbf{C}\|_{\text{F}} \leq \rho.$$

224 We deduce that $\mathbf{d}' \in \text{dom } \varphi$ and, consequently $\mathbf{C}' \in \text{dom } f$. This shows that $\mathbf{C} \in$
225 $\text{int}(\text{dom } f)$. \square

226 As an offspring of Theorem 2.1, we then get:

227 COROLLARY 2.6. Let f and g_0 be functions satisfying (4) and (5), respectively,
228 where $\varphi \in \Gamma_0(\mathbb{R}^n)$ is a Legendre function, $\psi \in \Gamma_0(\mathbb{R}^n)$, $\text{int}(\text{dom } \varphi) \cap \text{dom } \psi \neq \emptyset$, and
229 either ψ is bounded from below or $\varphi + \psi$ is supercoercive. Then, the D^f -proximity
230 operator of g_0 is defined at every $\mathbf{Y} \in \mathcal{S}_n$ such that $\mathbf{Y} = \mathbf{U}_{\mathbf{Y}} \text{Diag}(\mathbf{y})\mathbf{U}_{\mathbf{Y}}^{\top}$ with $\mathbf{U}_{\mathbf{Y}} \in$
231 \mathcal{O}_n and $\mathbf{y} \in \text{int}(\text{dom } \varphi)$, and it is expressed as

$$232 \quad (24) \quad \text{prox}_{g_0}^f(\mathbf{Y}) = \mathbf{U}_{\mathbf{Y}} \text{Diag}(\text{prox}_{\psi}^{\varphi}(\mathbf{y}))\mathbf{U}_{\mathbf{Y}}^{\top}.$$

234 *Proof.* According to the properties of spectral functions [38, Corollary 2.7],

235 (25) $\varphi \in \Gamma_0(\mathbb{R}^n)$ (resp. $\psi \in \Gamma_0(\mathbb{R}^n)$) $\Rightarrow f \in \Gamma_0(\mathcal{S}_n)$ (resp. $g_0 \in \Gamma_0(\mathcal{S}_n)$).

236 In addition, according to [38, Corollaries 3.3&3.5], since φ is a Legendre function,
 237 f is a Legendre function. It is also straightforward to check that, when ψ is lower
 238 bounded, then g_0 is lower bounded and, when $\varphi + \psi$ is supercoercive, then $f + g_0$
 239 is supercoercive. It also follows from Lemma 2.5 that $\text{int}(\text{dom } \varphi) \cap \text{dom } \psi \neq \emptyset \Leftrightarrow$
 240 $\text{int}(\text{dom } f) \cap \text{dom } g_0 \neq \emptyset$.

241 The above results show that the D^f -proximity operator of g_0 is properly defined
 242 as follows:

243 (26) $\text{prox}_{g_0}^f : \text{int}(\text{dom } f) \rightarrow \text{int}(\text{dom } f) \cap \text{dom } g_0$
 244 $\mathbf{Y} \mapsto \underset{\mathbf{C} \in \mathcal{S}_n}{\text{argmin}} g_0(\mathbf{C}) + D^f(\mathbf{C}, \mathbf{Y}).$
 245

246 This implies that computing the D^f -proximity operator of g_0 at $\mathbf{Y} \in \text{int}(\text{dom } f)$
 247 amounts to finding the unique solution to Problem (7) where $\mathbf{T} = \nabla f(\mathbf{Y})$. Let $\mathbf{Y} =$
 248 $\mathbf{U}_{\mathbf{Y}} \text{Diag}(\mathbf{y}) \mathbf{U}_{\mathbf{Y}}^\top$ with $\mathbf{U}_{\mathbf{Y}} \in \mathcal{O}_n$ and $\mathbf{y} \in \mathbb{R}^n$. By Lemma 2.5(ii), $\mathbf{Y} \in \text{int}(\text{dom } f) \Leftrightarrow$
 249 $\mathbf{y} \in \text{int}(\text{dom}(\varphi))$ and, according to [38, Corollary 3.3], $\mathbf{T} = \mathbf{U}_{\mathbf{Y}} \text{Diag}(\mathbf{t}) \mathbf{U}_{\mathbf{Y}}^\top$ with
 250 $\mathbf{t} = \nabla \varphi(\mathbf{y})$.

251 Furthermore, as φ is essentially strictly convex, it follows from [4, Theorem 5.9(ii)]
 252 that $\mathbf{t} = \nabla \varphi(\mathbf{y}) \in \text{int}(\text{dom } f^*)$, which according to [6, Theorem 14.17] is equivalent
 253 to the fact that $\mathbf{d} \mapsto \varphi(\mathbf{d}) - \mathbf{d}^\top \mathbf{t}$ is coercive. So, if ψ is lower-bounded, $\mathbf{d} \mapsto \varphi(\mathbf{d}) -$
 254 $\mathbf{d}^\top \mathbf{t} + \psi(\mathbf{d})$ is coercive. The same conclusion obviously holds if $\varphi + \psi$ is supercoercive.
 255 This shows that the assumptions of Theorem 2.1 are met. Consequently, applying
 256 this theorem yields

257 (27) $\text{prox}_{g_0}^f(\mathbf{Y}) = \mathbf{U}_{\mathbf{Y}} \text{Diag}(\widehat{\mathbf{d}}) \mathbf{U}_{\mathbf{Y}}^\top,$

258 where $\widehat{\mathbf{d}}$ minimizes

259 (28) $\mathbf{d} \mapsto \varphi(\mathbf{d}) - \mathbf{d}^\top \mathbf{t} + \psi(\mathbf{d})$

260 or, equivalently,

261 (29) $\mathbf{d} \mapsto \psi(\mathbf{d}) + D^\varphi(\mathbf{d}, \mathbf{y}).$

262 This shows that $\widehat{\mathbf{d}} = \text{prox}_{\psi}^\varphi(\mathbf{y})$. □

263 *Remark 2.7.* Corollary 2.6 extends known results concerning the case when $f =$
 264 $\|\cdot\|_{\mathbb{F}}/2$ [16]. A rigorous derivation of the proximity operator of spectral functions
 265 in $\Gamma_0(\mathcal{S}_n)$ for the standard Frobenius metric can be found in [6, Corollary 24.65].
 266 Our proof allows us to recover a similar result by adopting a more general approach.
 267 In particular, it is worth noticing that Theorem 2.1 does not require any convexity
 268 assumption.

269 **3. Proximal Iterative Approach.** Let us now turn to the more general case
 270 of the resolution of Problem (1) when $f \in \Gamma_0(\mathcal{S}_n)$ and $g_1 \neq 0$. Proximal splitting
 271 approaches for finding a minimizer of a sum of non-necessarily smooth functions have
 272 attracted a large interest in the last years [24, 51, 37, 15]. In these methods, the
 273 functions can be dealt with either via their gradient or their proximity operator de-
 274 pending on their differentiability properties. In this section, we first list a number of

275 proximity operators of scaled versions of $f - \text{trace}(\mathbf{T} \cdot) + g_0$, where f and g_0 , satisfy-
 276 ing (4) and (5), are chosen among several options that can be useful in a wide range
 277 of practical scenarios. Based on these results, we then propose a proximal splitting
 278 Douglas-Rachford algorithm to solve Problem (1).

279 **3.1. Proximity Operators.** By definition, computing the proximity operator
 280 of $\gamma(f - \text{trace}(\mathbf{T} \cdot) + g_0)$ with $\gamma \in]0, +\infty[$ at $\bar{\mathbf{C}} \in \mathcal{S}_n$ amounts to find a minimizer of
 281 the function

$$282 \quad (30) \quad \mathbf{C} \mapsto f(\mathbf{C}) - \text{trace}(\mathbf{T}\mathbf{C}) + g_0(\mathbf{C}) + \frac{1}{2\gamma} \|\mathbf{C} - \bar{\mathbf{C}}\|_{\mathbb{F}}^2$$

283 over \mathcal{S}_n . The (possibly empty) set of such minimizers is denoted by
 284 $\text{Prox}_{\gamma(f - \text{trace}(\mathbf{T} \cdot) + g_0)}(\bar{\mathbf{C}})$. As pointed out in Section 2, if $f + g_0 \in \Gamma_0(\mathcal{S}_n)$ then this
 285 set is a singleton $\{\text{prox}_{\gamma(f - \text{trace}(\mathbf{T} \cdot) + g_0)}(\bar{\mathbf{C}})\}$. We have the following characterization
 286 of this proximity operator:

287 **PROPOSITION 3.1.** *Let $\gamma \in]0, +\infty[$ and $\bar{\mathbf{C}} \in \mathcal{S}_n$. Let f and g_0 be functions sat-
 288 isfying (4) and (5), respectively, where $\varphi \in \Gamma_0(\mathbb{R}^n)$ and ψ is a lower-semicontinuous
 289 function such that $\text{dom } \varphi \cap \text{dom } \psi \neq \emptyset$. Let $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\mathbf{U} \in \mathcal{O}_n$ be such that
 290 $\bar{\mathbf{C}} + \gamma\mathbf{T} = \mathbf{U} \text{Diag}(\boldsymbol{\lambda}) \mathbf{U}^\top$.*

291 (i) *If ψ is lower bounded by an affine function then $\text{Prox}_{\gamma(\varphi + \psi)}(\boldsymbol{\lambda}) \neq \emptyset$ and, for
 292 every $\hat{\boldsymbol{\lambda}} \in \text{Prox}_{\gamma(\varphi + \psi)}(\boldsymbol{\lambda})$,*

$$293 \quad (31) \quad \mathbf{U} \text{Diag}(\hat{\boldsymbol{\lambda}}) \mathbf{U}^\top \in \text{Prox}_{\gamma(f - \text{trace}(\mathbf{T} \cdot) + g_0)}(\bar{\mathbf{C}}).$$

294 (ii) *If ψ is convex, then*

$$295 \quad (32) \quad \text{prox}_{\gamma(f - \text{trace}(\mathbf{T} \cdot) + g_0)}(\bar{\mathbf{C}}) = \mathbf{U} \text{Diag}(\text{prox}_{\gamma(\varphi + \psi)}(\boldsymbol{\lambda})) \mathbf{U}^\top.$$

296 *Proof.* (i): Since it has been assumed that f and g_0 are spectral functions, we
 297 have

$$298 \quad (33) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad f(\mathbf{C}) + g_0(\mathbf{C}) = \varphi(\mathbf{d}) + \psi(\mathbf{d}),$$

299 where $\mathbf{d} \in \mathbb{R}^n$ is a vector of the eigenvalues of \mathbf{C} . It can be noticed that minimizing
 300 (30) is obviously equivalent to minimize $\tilde{f} - \gamma^{-1} \text{trace}((\bar{\mathbf{C}} + \gamma\mathbf{T}) \cdot) + g_0$ where $\tilde{f} =$
 301 $f + \|\cdot\|_{\mathbb{F}}^2/(2\gamma)$. Then

$$302 \quad (34) \quad \tilde{f}(\mathbf{C}) = \tilde{\varphi}(\mathbf{d}),$$

303 where $\tilde{\varphi} = \varphi + \|\cdot\|_{\mathbb{F}}^2/(2\gamma)$. Since we have assumed that $\varphi \in \Gamma_0(\mathbb{R}^n)$, $\tilde{\varphi}$ is proper, lower-
 304 semicontinuous, and strongly convex. As ψ is lower bounded by an affine function, it
 305 follows that

$$306 \quad (35) \quad \mathbf{d} \mapsto \tilde{\varphi}(\mathbf{d}) - \gamma^{-1} \boldsymbol{\lambda}^\top \mathbf{d} + \psi(\mathbf{d})$$

307 is lower bounded by a strongly convex function and it is thus coercive. In addition,
 308 $\text{dom } \tilde{\varphi} = \text{dom } \varphi$, hence $\text{dom } \tilde{\varphi} \cap \text{dom } \psi \neq \emptyset$. Let us now apply [Theorem 2.1](#). Let $\hat{\boldsymbol{\lambda}}$ be
 309 a minimizer of (35). It can be claimed that $\hat{\mathbf{C}} = \mathbf{U} \text{Diag}(\hat{\boldsymbol{\lambda}}) \mathbf{U}^\top$ is a minimizer of (30).
 310 On the other hand, minimizing (35) is equivalent to minimize $\gamma(\varphi + \psi) + \frac{1}{2} \|\cdot - \boldsymbol{\lambda}\|^2$,
 311 which shows that $\hat{\boldsymbol{\lambda}} \in \text{Prox}_{\gamma(\varphi + \psi)}(\boldsymbol{\lambda})$.

312 (ii): If $\psi \in \Gamma_0(\mathbb{R}^n)$, then it is lower bounded by an affine function [6, Theo-
 313 rem 9.20]. Furthermore, $\varphi + \psi \in \Gamma_0(\mathbb{R}^n)$ and the proximity operator of $\gamma(\varphi + \psi)$ is
 314 thus single valued. On the other hand, we also have $\gamma(f - \text{trace}(\mathbf{T} \cdot) + g_0) \in \Gamma_0(\mathcal{S}_n)$
 315 [38, Corollary 2.7], and the proximity operator of this function is single valued too.
 316 The result directly follows from (i). \square

317 We will next focus on the use of Proposition 3.1 for three choices for f , namely the
 318 classical squared Frobenius norm, the minus log det functional, and the Von Neumann
 319 entropy, each choice being coupled with various possible choices for g_0 .

320 **3.1.1. Squared Frobenius Norm.** A suitable choice in Problem (1) is $f =$
 321 $\|\cdot\|_{\mathbb{F}}^2/2$ [72, 54, 19]. The squared Frobenius norm is the spectral function associated
 322 with the function $\varphi = \|\cdot\|^2/2$. It is worth mentioning that this choice for f allows us
 323 to rewrite the original Problem (1) under the form (3), where

$$324 \quad (36) \quad (\forall (\mathbf{C}, \mathbf{Y}) \in \mathcal{S}_n^2) \quad D^f(\mathbf{C}, \mathbf{Y}) = \frac{1}{2} \|\mathbf{C} - \mathbf{Y}\|_{\mathbb{F}}^2.$$

325 We have thus re-expressed Problem (1) as the determination of a proximal point of
 326 function g at \mathbf{T} in the Frobenius metric.

327 Table 1 presents several examples of spectral functions g_0 and the expression of the
 328 proximity operator of $\gamma(\varphi + \psi)$ with $\gamma \in]0, +\infty[$. These expressions were established
 329 by using the properties of proximity operators of functions defined on \mathbb{R}^n (see [20,
 330 Example 4.4] and [24, Tables 10.1 and 10.2]).
 331

332 *Remark 3.2.* Another option for g_0 is to choose it equal to $\mu\|\cdot\|_{\mathbb{S}}$ where $\mu \in$
 333 $]0, +\infty[$. For every $\gamma \in]0, +\infty[$, we have then

$$334 \quad (37) \quad (\forall \boldsymbol{\lambda} \in \mathbb{R}^n) \quad \text{prox}_{\gamma(\varphi+\psi)}(\boldsymbol{\lambda}) = \text{prox}_{\frac{\mu\gamma}{1+\gamma}\|\cdot\|_{+\infty}}\left(\frac{\boldsymbol{\lambda}}{1+\gamma}\right),$$

335 where $\|\cdot\|_{+\infty}$ is the infinity norm of \mathbb{R}^n . By noticing that $\|\cdot\|_{+\infty}$ is the conjugate
 336 function of the indicator function of B_{ℓ^1} , the unit ℓ^1 ball centered at 0 of \mathbb{R}^n , and
 337 using Moreau's decomposition formula, [6, Proposition 24.8(ix)] yields

$$338 \quad (38) \quad (\forall \boldsymbol{\lambda} \in \mathbb{R}^n) \quad \text{prox}_{\gamma(\varphi+\psi)}(\boldsymbol{\lambda}) = \frac{1}{1+\gamma} \left(\boldsymbol{\lambda} - \mu\gamma \text{proj}_{B_{\ell^1}} \left(\frac{\boldsymbol{\lambda}}{\mu\gamma} \right) \right).$$

339 The required projection onto B_{ℓ^1} can be computed through efficient algorithms [61,
 340 25].

341 **3.1.2. Logdet Function.** Another popular choice for f is the negative logarithmic
 342 determinant function [30, 58, 44, 48, 3, 31, 67, 18], which is defined as follows

$$343 \quad (39) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad f(\mathbf{C}) = \begin{cases} -\log \det(\mathbf{C}) & \text{if } \mathbf{C} \in \mathcal{S}_n^{++} \\ +\infty & \text{otherwise.} \end{cases}$$

344 The above function satisfies property (5) with

$$345 \quad (40) \quad (\forall \boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad \varphi(\boldsymbol{\lambda}) = \begin{cases} -\sum_{i=1}^n \log(\lambda_i) & \text{if } \boldsymbol{\lambda} \in]0, +\infty[^n \\ +\infty & \text{otherwise.} \end{cases}$$

TABLE 1

Proximity operators of $\gamma(\frac{1}{2}\|\cdot\|_F^2 + g_0)$ with $\gamma > 0$ evaluated at symmetric matrix with vector of eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n}$. For the inverse Schatten penalty, the function is set to $+\infty$ when the argument \mathbf{C} is not positive definite. E_1 denotes the set of matrices in \mathcal{S}_n with Frobenius norm less than or equal to α and E_2 the set of matrices in \mathcal{S}_n with eigenvalues between α and β . In the last line, the i -th component of the proximity operator is obtained by searching among the nonnegative roots of a third order polynomial those minimizing $\lambda'_i \mapsto \frac{1}{2}(\lambda'_i - |\lambda_i|)^2 + \gamma(\frac{1}{2}(\lambda'_i)^2 + \mu \log((\lambda'_i)^2 + \varepsilon))$.

$g_0(\mathbf{C}), \mu > 0$	$\text{prox}_{\gamma(\varphi+\psi)}(\boldsymbol{\lambda})$
Nuclear norm $\mu\mathcal{R}_1(\mathbf{C})$	$\left(\text{soft}_{\frac{\mu\gamma}{\gamma+1}}\left(\frac{\lambda_i}{\gamma+1}\right)\right)_{1 \leq i \leq n}$
Frobenius norm $\mu\ \mathbf{C}\ _F$	$\left(1 - \frac{\gamma\mu}{\ \boldsymbol{\lambda}\ }\right) \frac{\boldsymbol{\lambda}}{1+\gamma}$ if $\ \boldsymbol{\lambda}\ > \gamma\mu$ and $\mathbf{0}$ otherwise
Squared Frobenius norm $\mu\ \mathbf{C}\ _F^2$	$\frac{\boldsymbol{\lambda}}{1 + \gamma(1 + 2\mu)}$
Schatten 3-penalty $\mu\mathcal{R}_3^3(\mathbf{C})$	$(6\gamma\mu)^{-1} \left(\text{sign}(\lambda_i) \sqrt{(\gamma+1)^2 + 12 \lambda_i \gamma\mu} - \gamma - 1\right)_{1 \leq i \leq n}$
Schatten 4-penalty $\mu\mathcal{R}_4^4(\mathbf{C})$	$(8\gamma\mu)^{-1/3} \left(\sqrt[3]{\lambda_i + \sqrt{\lambda_i^2 + \zeta}} + \sqrt[3]{\lambda_i - \sqrt{\lambda_i^2 + \zeta}}\right)_{1 \leq i \leq n}$ with $\zeta = \frac{(\gamma+1)^3}{27\gamma\mu}$
Schatten 4/3-penalty $\mu\mathcal{R}_{4/3}^{4/3}(\mathbf{C})$	$\frac{1}{1+\gamma} \left(\lambda_i + \frac{4\gamma\mu}{3\sqrt[3]{2(1+\gamma)}} \left(\sqrt[3]{\sqrt{\lambda_i^2 + \zeta} - \lambda_i} - \sqrt[3]{\sqrt{\lambda_i^2 + \zeta} + \lambda_i}\right)\right)_{1 \leq i \leq n}$ with $\zeta = \frac{256(\gamma\mu)^3}{729(1+\gamma)}$
Schatten 3/2-penalty $\mu\mathcal{R}_{3/2}^{3/2}(\mathbf{C})$	$\frac{1}{1+\gamma} \left(\lambda_i + \frac{9\gamma^2\mu^2}{8(1+\gamma)} \text{sign}(\lambda_i) \left(1 - \sqrt{1 + \frac{16(1+\gamma)}{9\gamma^2\mu^2} \lambda_i }\right)\right)_{1 \leq i \leq n}$
Schatten p -penalty $\mu\mathcal{R}_p^p(\mathbf{C}), p \geq 1$	$(\text{sign}(\lambda_i)d_i)_{1 \leq i \leq n}$ with $(\forall i \in \{1, \dots, n\}) d_i \geq 0$ and $\mu\gamma p d_i^{p-1} + (\gamma+1)d_i = \lambda_i$
Inverse Schatten p -penalty $\mu\mathcal{R}_p^p(\mathbf{C}^{-1}), p > 0$	$(d_i)_{1 \leq i \leq n}$ with $(\forall i \in \{1, \dots, n\}) d_i > 0$ and $(\gamma+1)d_i^{p+2} - \lambda_i d_i^{p+1} = \mu\gamma p$
Bound on the Frobenius norm $\iota_{E_1}(\mathbf{C})$	$\alpha \frac{\boldsymbol{\lambda}}{\ \boldsymbol{\lambda}\ }$ if $\ \boldsymbol{\lambda}\ > \alpha(1+\gamma)$ and $\frac{\boldsymbol{\lambda}}{1+\gamma}$ otherwise, $\alpha \in [0, +\infty[$
Bounds on eigenvalues $\iota_{E_2}(\mathbf{C})$	$(\min(\max(\lambda_i/(\gamma+1), \alpha), \beta))_{1 \leq i \leq n}, [\alpha, \beta] \subset [-\infty, +\infty]$
Rank $\mu \text{rank}(\mathbf{C})$	$\left(\text{hard}_{\sqrt{\frac{2\mu\gamma}{1+\gamma}}}\left(\frac{\lambda_i}{1+\gamma}\right)\right)_{1 \leq i \leq n}$
Cauchy $\mu \log \det(\mathbf{C}^2 + \varepsilon \mathbf{I}_d), \varepsilon > 0$	$\in \{(\text{sign}(\lambda_i)d_i)_{1 \leq i \leq n} \mid (\forall i \in \{1, \dots, n\}) d_i \geq 0 \text{ and } (\gamma+1)d_i^3 - \lambda_i d_i^2 + (2\gamma\mu + \varepsilon(\gamma+1))d_i = \lambda_i \varepsilon\}$

346 Actually, for a given positive definite matrix, the value of function (39) simply reduces
347 to the Burg entropy of its eigenvalues. Hereagain, if $\mathbf{Y} \in \mathcal{S}_n^{++}$ and $\mathbf{T} = -\mathbf{Y}^{-1}$, we
348 can rewrite Problem (1) under the form (3), so that it becomes equivalent to the
349 computation of the proximity operator of g with respect to the Bregman divergence
350 given by

$$351 \quad (41) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad D^f(\mathbf{C}, \mathbf{Y}) = \begin{cases} \log\left(\frac{\det(\mathbf{Y})}{\det(\mathbf{C})}\right) + \text{trace}(\mathbf{Y}^{-1}\mathbf{C}) - n & \text{if } \mathbf{C} \in \mathcal{S}_n^{++} \\ +\infty & \text{otherwise.} \end{cases}$$

352 In Table 2, we list some particular choices for g_0 , and provide the associated
353 closed form expression of the proximity operator $\text{prox}_{\gamma(\varphi+\psi)}$ for $\gamma \in]0, +\infty[$, where φ
354 is defined in (40). These expressions were derived from [24, Table 10.2].

355 *Remark 3.3.* Let g_0 be any of the convex spectral functions listed in Table 2. Let
 356 \mathbf{W} be an invertible matrix in $\mathbb{R}^{n \times n}$, and let $\bar{\mathbf{C}} \in \mathcal{S}_n$. From the above results, one can
 357 deduce the minimizer of $\mathbf{C} \mapsto \gamma(f(\mathbf{C}) + g_0(\mathbf{W}\mathbf{C}\mathbf{W}^\top)) + \frac{1}{2}\|\mathbf{W}\mathbf{C}\mathbf{W}^\top - \bar{\mathbf{C}}\|_{\mathbb{F}}^2$ where
 358 $\gamma \in]0, +\infty[$. Indeed, by making a change of variable and by using basic properties of
 359 the log det function, this minimizer is equal to $\mathbf{W}^{-1} \text{prox}_{\gamma(f+g_0)}(\bar{\mathbf{C}})(\mathbf{W}^{-1})^\top$.

TABLE 2

Proximity operators of $\gamma(f + g_0)$ with $\gamma > 0$ and f given by (39), evaluated at a symmetric matrix with vector of eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n}$. For the inverse Schatten penalty, the function is set to $+\infty$ when the argument \mathbf{C} is not positive definite. E_2 denotes the set of matrices in \mathcal{S}_n with eigenvalues between α and β . In the last line, the i -th component of the proximity operator is obtained by searching among the positive roots of a fourth order polynomial those minimizing $\lambda'_i \mapsto \frac{1}{2}(\lambda'_i - \lambda_i)^2 + \gamma(\mu \log((\lambda'_i)^2 + \varepsilon) - \log \lambda'_i)$.

$g_0(\mathbf{C}), \mu > 0$	$\text{prox}_{\gamma(\varphi+\psi)}(\boldsymbol{\lambda})$
Nuclear norm $\mu \mathcal{R}_1(\mathbf{C})$	$\frac{1}{2} \left(\lambda_i - \gamma\mu + \sqrt{(\lambda_i - \gamma\mu)^2 + 4\gamma} \right)_{1 \leq i \leq n}$
Squared Frobenius norm $\mu \ \mathbf{C}\ _{\mathbb{F}}^2$	$\frac{1}{2(2\gamma\mu + 1)} \left(\lambda_i + \sqrt{\lambda_i^2 + 4\gamma(2\gamma\mu + 1)} \right)_{1 \leq i \leq n}$
Schatten p -penalty $\mu \mathcal{R}_p^p(\mathbf{C}), p \geq 1$	$(d_i)_{1 \leq i \leq n}$ with $(\forall i \in \{1, \dots, n\}) d_i > 0$ and $\mu\gamma p d_i^p + d_i^2 - \lambda_i d_i = \gamma$
Inverse Schatten p -penalty $\mu \mathcal{R}_p^p(\mathbf{C}^{-1}), p > 0$	$(d_i)_{1 \leq i \leq n}$ with $(\forall i \in \{1, \dots, n\}) d_i > 0$ and $d_i^{p+2} - \lambda_i d_i^{p+1} - \gamma d_i^p = \mu\gamma p$
Bounds on eigenvalues $\iota_{E_2}(\mathbf{C})$	$\left(\min \left(\max \left(\frac{1}{2}(\lambda_i + \sqrt{\lambda_i^2 + 4\gamma}), \alpha \right), \beta \right) \right)_{1 \leq i \leq n}, [\alpha, \beta] \subset [0, +\infty]$
Cauchy $\mu \log \det(\mathbf{C}^2 + \varepsilon \mathbf{I}_d), \varepsilon > 0$	$\in \left\{ (d_i)_{1 \leq i \leq n} \mid (\forall i \in \{1, \dots, n\}) d_i > 0 \text{ and } d_i^4 - \lambda d_i^3 + (\varepsilon + \gamma(2\mu - 1))d_i^2 - \varepsilon \lambda_i d_i = \gamma\varepsilon \right\}$

360 **3.1.3. Von Neumann Entropy.** Our third example is the negative Von Neu-
 361 mann entropy, which appears to be useful in some quantum mechanics problems [10].
 362 It is defined as

363 (42) $(\forall \mathbf{C} \in \mathcal{S}_n) \quad f(\mathbf{C}) = \begin{cases} \text{trace}(\mathbf{C} \log(\mathbf{C})) & \text{if } \mathbf{C} \in \mathcal{S}_n^+ \\ +\infty & \text{otherwise.} \end{cases}$

364 In the above expression, if $\mathbf{C} = \mathbf{U} \text{Diag}(\boldsymbol{\lambda}) \mathbf{U}^\top$ with $\boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n} \in]0, +\infty[^n$ and
 365 $\mathbf{U} \in \mathcal{O}_n$, then $\log(\mathbf{C}) = \mathbf{U} \text{Diag}((\log \lambda_i)_{1 \leq i \leq n}) \mathbf{U}^\top$. The logarithm of a symmetric
 366 definite positive matrix is uniquely defined and the function $\mathbf{C} \mapsto \mathbf{C} \log(\mathbf{C})$ can be
 367 extended by continuity on \mathcal{S}_n^+ similarly to the case when $n = 1$. Thus, f is the spectral
 368 function associated with

369 (43) $(\forall \boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad \varphi(\boldsymbol{\lambda}) = \begin{cases} \sum_{i=1}^n \lambda_i \log(\lambda_i) & \text{if } \boldsymbol{\lambda} \in [0, +\infty[^n \\ +\infty & \text{otherwise.} \end{cases}$

370 Note that the Von Neumann entropy defined for symmetric matrices is simply equal
 371 to the well-known Shannon entropy [27] of the input eigenvalues. With this choice
 372 for function f , by setting $\mathbf{T} = \log(\mathbf{Y}) + \mathbf{I}_d$ where $\mathbf{Y} \in \mathcal{S}_n^{++}$, Problem (1) can be
 373 recast under the form (3), so that it becomes equivalent to the computation of the

374 proximity operator of g with respect to the Bregman divergence associated with the
 375 Von Neumann entropy:

376

377

$$(\forall \mathbf{C} \in \mathcal{S}_n) \quad D^f(\mathbf{C}, \mathbf{Y}) = \begin{cases} \text{trace}(\mathbf{C} \log(\mathbf{C}) - \mathbf{Y} \log(\mathbf{Y}) - (\log(\mathbf{Y}) + \text{Id})(\mathbf{C} - \mathbf{Y})) & \text{if } \mathbf{C} \in \mathcal{S}_n^+ \\ +\infty & \text{otherwise.} \end{cases}$$

378

379

380 We provide in Table 3 a list of closed form expressions of the proximity operator
 381 of $\gamma(f + g_0)$ for several choices of the spectral function g_0 .

TABLE 3

Proximity operators of $\gamma(f+g_0)$ with $\gamma > 0$ and f given by (42), evaluated at a symmetric matrix with vector of eigenvalues $\boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n}$. E_2 denotes the set of matrices in \mathcal{S}_n with eigenvalues between α and β . $W(\cdot)$ denotes the W-Lambert function [26].

$g_0(\mathbf{C}), \mu > 0$	$\text{prox}_{\gamma(\varphi+\psi)}(\boldsymbol{\lambda})$
Nuclear norm $\mu \mathcal{R}_1(\mathbf{C})$	$\gamma \left(W \left(\frac{1}{\gamma} \exp \left(\frac{\lambda_i}{\gamma} - \mu - 1 \right) \right) \right)_{1 \leq i \leq n}$
Squared Frobenius norm $\mu \ \mathbf{C}\ _{\mathbb{F}}^2$	$\frac{\gamma}{2\mu\gamma+1} \left(W \left(\frac{2\mu\gamma+1}{\gamma} \exp \left(\frac{\lambda_i}{\gamma} - 1 \right) \right) \right)_{1 \leq i \leq n}$
Schatten p -penalty $\mu \mathcal{R}_p(\mathbf{C}), p \geq 1$	$(d_i)_{1 \leq i \leq n}$ with $(\forall i \in \{1, \dots, n\}) d_i > 0$ and $p\mu\gamma d_i^{p-1} + d_i + \gamma \log d_i + \gamma = \lambda_i$
Bounds on eigenvalues $\iota_{E_2}(\mathbf{C})$	$\left(\min \left(\max \left(\gamma W \left(\frac{1}{\gamma} \exp \left(\frac{\lambda_i}{\gamma} - 1 \right) \right), \alpha \right), \beta \right) \right)_{1 \leq i \leq n}, [\alpha, \beta] \subset [0, +\infty]$
Rank	$(d_i)_{1 \leq i \leq n}$ with
$\mu \text{rank}(\mathbf{C})$	$(\forall i \in \{1, \dots, n\}) d_i = \begin{cases} \rho_i & \text{if } \rho_i > \chi \\ 0 \text{ or } \rho_i & \text{if } \rho_i = \chi \\ 0 & \text{otherwise} \end{cases} \text{ and } \begin{cases} \chi = \sqrt{\gamma(\gamma + 2\mu)} - \gamma, \\ \rho_i = \gamma W \left(\frac{1}{\gamma} \exp \left(\frac{\lambda_i}{\gamma} - 1 \right) \right) \end{cases}$

382

383

384

385

386

387

388

389

390

391

392

393

3.2. Douglas-Rachford Algorithm. We now propose a Douglas-Rachford (DR) approach ([41, 24, 23]) for numerically solving Problem (1). The DR method minimizes the sum of $f - \text{trace}(\mathbf{T}\cdot) + g_0$ and g_1 by alternately computing proximity operators of each of these functions. Proposition 3.1 allows us to calculate the proximity operator of $\gamma(f - \text{trace}(\mathbf{T}\cdot) + g_0)$ with $\gamma \in]0, +\infty[$, by possibly using the expressions listed in Tables 1, 2, and 3. Since g_1 is not a spectral function, $\text{prox}_{\gamma g_1}$ has to be derived from other expressions of proximity operators. For instance, if g_1 is a separable sum of functions of its elements, e.g. $g = \|\cdot\|_1$, standard expressions for the proximity operator of vector functions can be employed [20, 24].¹

The computations to be performed are summarized in Algorithm 1. We state a convergence theorem in the matrix framework, which is an offspring of existing results in arbitrary Hilbert spaces (see, for example, [24] and [52, Proposition 3.5]).

394

395

396

397

398

399

THEOREM 3.4. *Let f and g_0 be functions satisfying (4) and (5), respectively, where $\varphi \in \Gamma_0(\mathbb{R}^n)$ and $\psi \in \Gamma_0(\mathbb{R}^n)$. Let $g_1 \in \Gamma_0(\mathcal{S}_n)$ be such that $f - \text{trace}(\mathbf{T}\cdot) + g_0 + g_1$ is coercive. Assume that the intersection of the relative interiors of the domains of $f + g_0$ and g_1 is non empty. Let $(\alpha^{(k)})_{k \geq 0}$ be a sequence in $[0, 2]$ such that $\sum_{k=0}^{+\infty} \alpha^{(k)}(2 - \alpha^{(k)}) = +\infty$. Then, the sequences $(\mathbf{C}^{(k+\frac{1}{2})})_{k \geq 0}$ and $(\text{prox}_{\gamma g_1}(2\mathbf{C}^{(k+\frac{1}{2})} - \mathbf{C}^{(k)}))_{k \geq 0}$ generated by Algorithm 1 converge to a solution to Problem (1) where $g = g_0 + g_1$.*

¹See also <http://proximity-operator.net>.

Algorithm 1 Douglas–Rachford Algorithm for solving Problem (1)

- 1: Let \mathbf{T} be a given matrix in \mathcal{S}_n , set $\gamma > 0$ and $\mathbf{C}^{(0)} \in \mathcal{S}_n$.
- 2: **for** $k = 0, 1, \dots$ **do**
- 3: Diagonalize $\mathbf{C}^{(k)} + \gamma\mathbf{T}$, i.e. find $\mathbf{U}^{(k)} \in \mathcal{O}_n$ and $\boldsymbol{\lambda}^{(k)} \in \mathbb{R}^n$ such that

$$\mathbf{C}^{(k)} + \gamma\mathbf{T} = \mathbf{U}^{(k)} \text{Diag}(\boldsymbol{\lambda}^{(k)}) (\mathbf{U}^{(k)})^\top$$

- 4: $\mathbf{d}^{(k+\frac{1}{2})} \in \text{Prox}_{\gamma(\varphi+\psi)}(\boldsymbol{\lambda}^{(k)})$
 - 5: $\mathbf{C}^{(k+\frac{1}{2})} = \mathbf{U}^{(k)} \text{Diag}(\mathbf{d}^{(k+\frac{1}{2})}) (\mathbf{U}^{(k)})^\top$
 - 6: Choose $\alpha^{(k)} \in [0, 2]$
 - 7: $\mathbf{C}^{(k+1)} \in \mathbf{C}^{(k)} + \alpha^{(k)} \left(\text{Prox}_{\gamma g_1}(2\mathbf{C}^{(k+\frac{1}{2})} - \mathbf{C}^{(k)}) - \mathbf{C}^{(k+\frac{1}{2})} \right)$.
 - 8: **end for**
-

400 We have restricted the above convergence analysis to the convex case. Note however
401 that recent convergence results for the DR algorithm in a non-convex setting are
402 available in [1, 39] for specific choices of the involved functionals.

403 **3.3. Positive Semi-Definite Constraint.** Instead of solving Problem (1), one
404 may be interested in:

$$405 \quad (44) \quad \underset{\mathbf{C} \in \mathcal{S}_n^+}{\text{minimize}} \quad f(\mathbf{C}) - \text{trace}(\mathbf{C}\mathbf{T}) + g(\mathbf{C}),$$

406 when $\text{dom } f \cap \text{dom } g \not\subset \mathcal{S}_n^+$. This problem can be recast as minimizing over \mathcal{S}_n
407 $f - \text{trace}(\cdot\mathbf{T}) + \tilde{g}_0 + g_1$ where $\tilde{g}_0 = g_0 + \iota_{\mathcal{S}_n^+}$. We are thus coming back to the original
408 formulation where \tilde{g}_0 has been substituted for g_0 . In order to solve this problem with
409 the proposed proximal approach, a useful result is stated below.

410 **PROPOSITION 3.5.** *Let $\gamma \in]0, +\infty[$ and $\bar{\mathbf{C}} \in \mathcal{S}_n$. Let f and g_0 be functions satis-*
411 *fying (4) and (5), respectively, where $\varphi \in \Gamma_0(\mathbb{R}^n)$ and $\psi \in \Gamma_0(\mathbb{R}^n)$. Assume that*

$$412 \quad (45) \quad (\forall \boldsymbol{\lambda}' = (\lambda'_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad \varphi(\boldsymbol{\lambda}') + \psi(\boldsymbol{\lambda}') = \sum_{i=1}^n \rho_i(\lambda'_i)$$

413 *where, for every $i \in \{1, \dots, n\}$, $\rho_i: \mathbb{R} \rightarrow]-\infty, +\infty]$ is such that $\text{dom } \rho_i \cap [0, +\infty[\neq \emptyset$.*
414 *Let $\boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n} \in \mathbb{R}^n$ and $\mathbf{U} \in \mathcal{O}_n$ be such that $\bar{\mathbf{C}} + \gamma\mathbf{T} = \mathbf{U} \text{Diag}(\boldsymbol{\lambda}) \mathbf{U}^\top$. Then*

$$415 \quad (46) \quad \text{prox}_{\gamma(f - \text{trace}(\mathbf{T}\cdot) + \tilde{g}_0)}(\bar{\mathbf{C}}) = \mathbf{U} \text{Diag} \left(\left(\max(0, \text{prox}_{\gamma\rho_i}(\lambda_i)) \right)_{1 \leq i \leq n} \right) \mathbf{U}^\top.$$

416 *Proof.* Expression (46) readily follows from Proposition 3.1(ii) and [21, Proposi-
417 tion 2.2]. \square

418 **4. Application to Covariance Matrix Estimation.** Estimating the covari-
419 ance matrix of a random vector is a key problem in statistics, signal processing over
420 graphs, and machine learning. Nonetheless, in existing optimization techniques, little
421 attention is usually paid to the presence of noise corrupting the available observations.
422 We show in this section how the results obtained in the previous sections can be used
423 to tackle this problem in various contexts.

424 **4.1. Model and Proposed Approaches.** Let $\mathbf{S} \in \mathcal{S}_n^+$ be a sample estimate of
 425 a covariance matrix Σ which is assumed to be decomposed as

$$426 \quad (47) \quad \Sigma = \mathbf{Y}^* + \sigma^2 \mathbf{I}_d$$

427 where $\sigma \in [0, +\infty[$ and $\mathbf{Y}^* \in \mathcal{S}_n^+$ may have a low-rank structure. Our objective in
 428 this section will be to propose variational methods to provide an estimate of \mathbf{Y}^* from
 429 \mathbf{S} by assuming that σ is known. Such a problem arises when considering the following
 430 observation model [59]:

$$431 \quad (48) \quad (\forall i \in \{1, \dots, N\}) \quad \mathbf{x}^{(i)} = \mathbf{A} \mathbf{s}^{(i)} + \mathbf{e}^{(i)}$$

432 where $\mathbf{A} \in \mathbb{R}^{n \times m}$ with $m \leq n$ and, for every $i \in \{1, \dots, N\}$, $\mathbf{s}^{(i)} \in \mathbb{R}^m$ and $\mathbf{e}^{(i)} \in \mathbb{R}^n$
 433 are realizations of mutually independent identically distributed Gaussian multivalued
 434 random variables with zero mean and covariance matrices $\mathbf{P} \in \mathcal{S}_m^{++}$ and $\sigma^2 \mathbf{I}_d$, re-
 435 spectively. This model has been employed for instance in [60, 63] in the context of
 436 the ‘‘Relevant Vector Machine problem’’. The covariance matrix Σ of the noisy input
 437 data $(\mathbf{x}^{(i)})_{1 \leq i \leq N}$ takes the form (47) with $\mathbf{Y}^* = \mathbf{A} \mathbf{P} \mathbf{A}^\top$. On the other hand, a simple
 438 estimate of Σ from the observed data $(\mathbf{x}^{(i)})_{1 \leq i \leq N}$ is

$$439 \quad (49) \quad \mathbf{S} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}^{(i)} (\mathbf{x}^{(i)})^\top.$$

440 *Covariance-based model.* A first estimate $\widehat{\mathbf{Y}}$ of \mathbf{Y}^* is given by

$$441 \quad (50) \quad \widehat{\mathbf{Y}} = \underset{\mathbf{Y} \in \mathcal{S}_n^+}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{S} + \sigma^2 \mathbf{I}_d\|_{\mathbb{F}}^2 + g_0(\mathbf{Y}) + g_1(\mathbf{Y}),$$

442 where \mathbf{S} is the empirical covariance matrix, g_0 satisfies (5) with $\psi \in \Gamma_0(\mathbb{R}^n)$, $g_1 \in$
 443 $\Gamma_0(\mathcal{S}_n)$, and the intersection of the relative interiors of the domains of g_0 and g_1
 444 is assumed to be non empty. A particular instance of this model with $\sigma = 0$, $g_0 =$
 445 $\mu_0 \mathcal{R}_1$, $g_1 = \mu_1 \|\cdot\|_1$, and $(\mu_0, \mu_1) \in [0, +\infty[^2$ was investigated in [72] and [54] for
 446 estimating sparse low-rank covariance matrices. In the latter reference, an application
 447 to real data processing arising from protein interaction and social network analysis
 448 is presented. One can observe that Problem (50) takes the form (44) by setting
 449 $f = \frac{1}{2} \|\cdot\|_{\mathbb{F}}^2$ and $\mathbf{T} = \mathbf{S} - \sigma^2 \mathbf{I}_d$. This allows us to solve (50) with Algorithm 1. Since it
 450 is assumed that g_0 satisfies (5), the proximity step on $f + g_0 + \iota_{\mathcal{S}_n^+}$ can be performed
 451 by employing Proposition 3.5 and formulas from Table 1. The resulting Douglas–
 452 Rachford procedure can thus be viewed as an alternative to the methods developed
 453 in [54] and [72]. Let us emphasize that these two algorithms were devised to solve an
 454 instance of (50) corresponding to the aforementioned specific choices for g_0 and g_1 ,
 455 while our approach leaves more freedom in the choice of the regularization functions.

456 *Precision-based model.* An alternative strategy consists of focusing on the esti-
 457 mation of the inverse of the covariance matrix, i.e. the *precision* matrix $\mathbf{C}^* = (\mathbf{Y}^*)^{-1}$
 458 by assuming that $\mathbf{Y}^* \in \mathcal{S}_n^{++}$ but may have very small eigenvalues in order to model
 459 a possible low-rank structure. Tackling the problem from this viewpoint leads us to
 460 propose the following penalized negative log-likelihood cost function:

$$461 \quad (51) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad \mathcal{F}(\mathbf{C}) = f(\mathbf{C}) + \mathcal{T}_{\mathbf{S}}(\mathbf{C}) + g_0(\mathbf{C}) + g_1(\mathbf{C})$$

462 where

$$463 \quad (52) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad f(\mathbf{C}) = \begin{cases} \log \det (\mathbf{C}^{-1} + \sigma^2 \mathbf{I}_d) & \text{if } \mathbf{C} \in \mathcal{S}_n^{++} \\ +\infty & \text{otherwise,} \end{cases}$$

$$464 \quad (53) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad \mathcal{T}_{\mathbf{S}}(\mathbf{C}) = \begin{cases} \text{trace} \left((\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1} \mathbf{C} \mathbf{S} \right) & \text{if } \mathbf{C} \in \mathcal{S}_n^+ \\ +\infty & \text{otherwise,} \end{cases}$$

466 $g_0 \in \Gamma_0(\mathcal{S}_n)$ satisfies (5) with $\psi \in \Gamma_0(\mathbb{R}^n)$, and $g_1 \in \Gamma_0(\mathcal{S}_n)$. Typical choices of
467 interest for the latter two functions are

$$468 \quad (54) \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad g_0(\mathbf{C}) = \begin{cases} \mu_0 \mathcal{R}_1(\mathbf{C}^{-1}) & \text{if } \mathbf{C} \in \mathcal{S}_n^{++} \\ +\infty & \text{otherwise,} \end{cases}$$

469 and $g_1 = \mu_1 \|\cdot\|_1$ with $(\mu_0, \mu_1) \in [0, +\infty]^2$. The first function serves to promote
470 a desired low-rank property by penalizing small eigenvalues of the precision matrix,
471 whereas the second one enforces the sparsity of this matrix as it is usual in graph
472 inference problems. This constitutes a main difference with respect to the covariance-
473 based model which is more suitable to estimate sparse covariance matrices. Note that
474 the standard Graphical Lasso framework [31] is then recovered by setting $\sigma = 0$ and
475 $\mu_0 = 0$. The advantage of our formulation is that it allows us to consider more flexible
476 variational models while accounting for the presence of noise corrupting the observed
477 data. The main difficulty however is that Algorithm 1 cannot be directly applied to
478 minimize \mathcal{F} . In Subsection 4.2, we will study in more details the properties of the
479 cost function. This will allow us to derive a novel optimization algorithm making use
480 of our previously developed Douglas-Rachford scheme for its inner steps

481 **4.2. Study of Objective Function \mathcal{F} .** The following lemma will reveal useful
482 in our subsequent analysis.

483 **LEMMA 4.1.** *Let $\sigma \in]0, +\infty[$. Let $h:]0, \sigma^{-2}[\rightarrow \mathbb{R}$ be a twice differentiable function
484 and let*

$$485 \quad (55) \quad u:]0, +\infty[\rightarrow \mathbb{R}: \lambda \mapsto \frac{\lambda}{1 + \sigma^2 \lambda}.$$

486 *The composition $h \circ u$ is convex on $]0, +\infty[$ if and only if*

$$487 \quad (56) \quad (\forall v \in]0, \sigma^{-2}[) \quad \ddot{h}(v)(1 - \sigma^2 v) - 2\sigma^2 \dot{h}(v) \geq 0,$$

488 *where \dot{h} (resp. \ddot{h}) denotes the first (resp. second) derivative of h .*

489 *Proof.* The result directly follows from the calculation of the second-order deriva-
490 tive of $h \circ u$. □

491 Let us now note that f is a spectral function fulfilling (4) with

$$492 \quad (57) \quad (\forall \boldsymbol{\lambda} = (\lambda_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad \varphi(\boldsymbol{\lambda}) = \begin{cases} -\sum_{i=1}^n \log(u(\lambda_i)) & \text{if } \boldsymbol{\lambda} \in]0, +\infty[^n \\ +\infty & \text{otherwise,} \end{cases}$$

493 where u is defined by (55). According to Lemma 4.1 (with $h = -\log$), $f \in \Gamma_0(\mathcal{S}_n)$.
494 Thus, the assumptions made on g_0 and g_1 , allow us to deduce that $f + g_0 + g_1$ is
495 convex and lower-semicontinuous on \mathcal{S}_n .

496 Let us now focus on the properties of the second term in (51).

497 LEMMA 4.2. Let $\mathbf{S} \in \mathcal{S}_n^+$. The function $\mathcal{T}_{\mathbf{S}}$ in (53) is concave on \mathcal{S}_n^+ .

498 *Proof.* By using differential calculus rules in [45], we will show that the Hessian
 499 of $-\mathcal{T}_{\mathbf{S}}$ evaluated at any matrix in \mathcal{S}_n^{++} is a positive semidefinite operator. In order
 500 to lighten our notation, for every invertible matrix \mathbf{C} , let us define $\mathbf{M} = \mathbf{C}^{-1} + \sigma^2 \mathbf{I}_d$.
 501 Then, the first-order differential of $\mathcal{T}_{\mathbf{S}}$ at every $\mathbf{C} \in \mathcal{S}_n^{++}$ is

$$\begin{aligned}
 502 \quad d \operatorname{trace}(\mathcal{T}_{\mathbf{S}}(\mathbf{C})) &= \operatorname{trace}((d\mathbf{M}^{-1})\mathbf{S}) \\
 503 &= \operatorname{trace}(-\mathbf{M}^{-1}(d\mathbf{M})\mathbf{M}^{-1}\mathbf{S}) \\
 504 &= \operatorname{trace}\left(\left(\mathbf{C}^{-1} + \sigma^2 \mathbf{I}_d\right)^{-1} \mathbf{S} \left(\mathbf{C}^{-1} + \sigma^2 \mathbf{I}_d\right)^{-1} \mathbf{C}^{-1}(d\mathbf{C})\mathbf{C}^{-1}\right) \\
 505 \quad (58) \quad &= \operatorname{trace}\left(\left(\mathbf{I}_d + \sigma^2 \mathbf{C}\right)^{-1} \mathbf{S} \left(\mathbf{I}_d + \sigma^2 \mathbf{C}\right)^{-1} (d\mathbf{C})\right).
 \end{aligned}$$

506 We have used the expression of the differential of the inverse [45, Chapter 8, Theo-
 507 rem 3] and the invariance of the trace with respect to cyclic permutations. It follows
 508 from (58) that the gradient of $\mathcal{T}_{\mathbf{S}}$ reads

$$509 \quad (59) \quad (\forall \mathbf{C} \in \mathcal{S}_n^{++}) \quad \nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) = \left(\mathbf{I}_d + \sigma^2 \mathbf{C}\right)^{-1} \mathbf{S} \left(\mathbf{I}_d + \sigma^2 \mathbf{C}\right)^{-1}.$$

510 In order to calculate the Hessian \mathfrak{H} of $\mathcal{T}_{\mathbf{S}}$, we calculate the differential of $\nabla \mathcal{T}_{\mathbf{S}}$. Again,
 511 in order to simplify our notation, for every matrix \mathbf{C} , we define

$$512 \quad (60) \quad \mathbf{N} = \mathbf{I}_d + \sigma^2 \mathbf{C} \quad \Rightarrow \quad d\mathbf{N} = \sigma^2 d\mathbf{C}.$$

513 The differential of $\nabla \mathcal{T}_{\mathbf{S}}$ at every $\mathbf{C} \in \mathcal{S}_n^{++}$ then reads

$$\begin{aligned}
 514 \quad d \operatorname{vect}(\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C})) &= \operatorname{vect}(d(\mathbf{N}^{-1} \mathbf{S} \mathbf{N}^{-1})) \\
 515 &= \operatorname{vect}((d\mathbf{N}^{-1})\mathbf{S}\mathbf{N}^{-1} + \mathbf{N}^{-1}(d\mathbf{S})\mathbf{N}^{-1}) \\
 516 &= -\operatorname{vect}(\mathbf{N}^{-1}(d\mathbf{N})\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}) - \operatorname{vect}(\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}(d\mathbf{N})\mathbf{N}^{-1}) \\
 517 &= -\left((\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1})^\top \otimes \mathbf{N}^{-1}\right) \operatorname{vect}(d\mathbf{N}) - \left((\mathbf{N}^{-1})^\top \otimes \mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}\right) \operatorname{vect}(d\mathbf{N}) \\
 518 &= -\left((\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1}) \otimes \mathbf{N}^{-1} + \mathbf{N}^{-1} \otimes (\mathbf{N}^{-1}\mathbf{S}\mathbf{N}^{-1})\right) d \operatorname{vect}(\mathbf{N}) \\
 519 &= \mathfrak{H}(\mathbf{C}) d \operatorname{vect}(\mathbf{C})
 \end{aligned}$$

520 with

$$521 \quad (61) \quad \mathfrak{H}(\mathbf{C}) = -\sigma^2 \left(\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \otimes (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1} + (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1} \otimes \nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \right).$$

522 To derive the above expression, we have used the facts that, for every $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\mathbf{X} \in$
 523 $\mathbb{R}^{m \times p}$, and $\mathbf{B} \in \mathbb{R}^{p \times q}$, $\operatorname{vect}(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{A}) \operatorname{vect} \mathbf{X}$ [45, Chapter 2, Theorem 2]
 524 and that matrices \mathbf{N} and \mathbf{S} are symmetric.

525 Let us now check that, for every $\mathbf{C} \in \mathcal{S}_n^{++}$, $\mathfrak{H}(\mathbf{C})$ is negative semidefinite. It
 526 follows from expression (59), the symmetry of \mathbf{C} , and the positive semidefiniteness of
 527 \mathbf{S} that $\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C})$ belongs to \mathcal{S}_n^+ . Since

$$\begin{aligned}
 528 \quad (\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \otimes (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1})^\top &= (\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}))^\top \otimes ((\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1})^\top \\
 529 &= \nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \otimes (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1}, \\
 530
 \end{aligned}$$

531 $\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \otimes (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1}$ is symmetric. Let us denote by $(\gamma_i)_{1 \leq i \leq n} \in [0, +\infty[^n$
 532 the eigenvalues of $\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C})$ and by $(\zeta_i)_{1 \leq i \leq n} \in [0, +\infty[^n$ those of \mathbf{C} . Accord-
 533 ing to [45, Chapter 2, Theorem 1], the eigenvalues of $\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \otimes (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1}$ are

534 $(\gamma_i/(1 + \sigma^2 \zeta_j))_{1 \leq i, j \leq n}$ and they are therefore nonnegative. This allows us to claim
535 that $\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \otimes (\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1}$ belongs to $\mathcal{S}_{n^2}^+$. For similar reasons, $(\mathbf{I}_d + \sigma^2 \mathbf{C})^{-1} \otimes$
536 $\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \in \mathcal{S}_{n^2}^+$, which allows us to conclude that $-\mathfrak{H}(\mathbf{C}) \in \mathcal{S}_{n^2}^+$. Hence, we have
537 proved that $\mathcal{T}_{\mathbf{S}}$ is concave on \mathcal{S}_n^{++} . By continuity of $\mathcal{T}_{\mathbf{S}}$ relative to \mathcal{S}_n^+ , the concavity
538 property extends on \mathcal{S}_n^+ . \square

539 As a last worth mentioning property, $\mathcal{T}_{\mathbf{S}}$ is bounded on \mathcal{S}_n^{++} . So, if $\text{dom } f \cap \text{dom } g_0 \cap$
540 $\text{dom } g_1 \neq \emptyset$ and $f + g_0 + g_1$ is coercive, then there exists a minimizer of \mathcal{F} . Because
541 of the form of f , the coercivity condition is satisfied if $g_0 + g_1$ is lower bounded and
542 $\lim_{\mathbf{C} \in \mathcal{S}_n^+, \|\mathbf{C}\| \rightarrow +\infty} g_0(\mathbf{C}) + g_1(\mathbf{C}) = +\infty$.

543 **4.3. Minimization Algorithm for \mathcal{F} .** In order to find a minimizer of \mathcal{F} , we
544 propose a *Majorize–Minimize* (MM) approach, following the ideas in [22, 59, 35, 36].
545 At each iteration of an MM algorithm, one constructs a tangent function that ma-
546 jorizes the given cost function and is equal to it at the current iterate. The next iterate
547 is obtained by minimizing this tangent majorant function, resulting in a sequence of
548 iterates that reduces the cost function value monotonically. According to the results
549 stated in the previous section, our objective function reads as a difference of convex
550 terms. We propose to build a majorizing approximation of function $\mathcal{T}_{\mathbf{S}}$ at $\mathbf{C}' \in \mathcal{S}_n^{++}$
551 by exploiting Lemma 4.2 and the classical concavity inequality on $\mathcal{T}_{\mathbf{S}}$:

$$552 \quad (62) \quad (\forall \mathbf{C} \in \mathcal{S}_n^{++}) \quad \mathcal{T}_{\mathbf{S}}(\mathbf{C}) \leq \mathcal{T}_{\mathbf{S}}(\mathbf{C}') + \text{trace}(\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}')(\mathbf{C} - \mathbf{C}')).$$

553 As f is finite only on \mathcal{S}_n^{++} , a tangent majorant of the cost function (51) at \mathbf{C}' reads:

$$554 \quad (\forall \mathbf{C} \in \mathcal{S}_n) \quad \mathcal{G}(\mathbf{C} \mid \mathbf{C}') = f(\mathbf{C}) + \mathcal{T}_{\mathbf{S}}(\mathbf{C}') + \text{trace}(\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}')(\mathbf{C} - \mathbf{C}')) + g_0(\mathbf{C}) + g_1(\mathbf{C}).$$

555 This leads to the general MM scheme:

$$556 \quad (63) \quad (\forall \ell \in \mathbb{N}) \quad \mathbf{C}^{(\ell+1)} \in \underset{\mathbf{C} \in \mathcal{S}_n}{\text{Argmin}} f(\mathbf{C}) + \text{trace}(\nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}^{(\ell)})\mathbf{C}) + g_0(\mathbf{C}) + g_1(\mathbf{C})$$

557 with $\mathbf{C}^{(0)} \in \mathcal{S}_n^{++}$. At each iteration of the MM algorithm, we have then to solve
558 a convex optimization problem of the form (1). In the case when $g_1 \equiv 0$, we can
559 employ the procedure described in Section 2 to perform this task in a direct manner.
560 The presence of a regularization term $g_1 \neq 0$ usually prevents us to have an explicit
561 solution to the inner minimization problem involved in the MM procedure. We then
562 propose in Algorithm 2 to resort to the Douglas–Rachford approach in Section 3 to
563 solve it iteratively.

Algorithm 2 MM algorithm with DR inner steps

-
- 1: Let $\mathbf{S} \in \mathcal{S}_n^+$ be the data matrix. Let φ be as in (57), let $\psi \in \Gamma_0(\mathbb{R}^n)$ be associated with g_0 . Let $(\gamma_\ell)_{\ell \in \mathbb{N}}$ be a sequence in $]0, +\infty[$. Set $\mathbf{C}^{(0,0)} = \mathbf{C}^{(0)} \in \mathcal{S}_n^{++}$.
 - 2: **for** $\ell = 0, 1, \dots$ **do**
 - 3: **for** $k = 0, 1, \dots$ **do**
 - 4: Compute $\mathbf{U}^{(\ell,k)} \in \mathcal{O}_n$ and $\boldsymbol{\lambda}^{(\ell,k)} \in \mathbb{R}^n$ such that

$$\mathbf{C}^{(\ell,k)} - \gamma_\ell \nabla \mathcal{T}_{\mathbf{S}}(\mathbf{C}^{(\ell)}) = \mathbf{U}^{(\ell,k)} \text{Diag}(\boldsymbol{\lambda}^{(\ell,k)}) \left(\mathbf{U}^{(\ell,k)} \right)^\top$$

- 5: $\mathbf{d}^{(\ell,k+\frac{1}{2})} = \text{prox}_{\gamma_\ell(\varphi+\psi)}(\boldsymbol{\lambda}^{(\ell,k)})$
 - 6: $\mathbf{C}^{(\ell,k+\frac{1}{2})} = \mathbf{U}^{(\ell,k)} \text{Diag}(\mathbf{d}^{(\ell,k+\frac{1}{2})}) \left(\mathbf{U}^{(\ell,k)} \right)^\top$
 - 7: **if** Convergence of MM sub-iteration is reached **then**
 - 8: $\mathbf{C}^{(\ell+1)} = \mathbf{C}^{(\ell,k+\frac{1}{2})}$
 - 9: $\mathbf{C}^{(\ell+1,0)} = \mathbf{C}^{(\ell,k)}$
 - 10: **exit** inner loop
 - 11: **end if**
 - 12: Choose $\alpha_{\ell,k} \in]0, 2[$
 - 13: $\mathbf{C}^{(\ell,k+1)} = \mathbf{C}^{(\ell,k)} + \alpha_{\ell,k} \left(\text{prox}_{\gamma_\ell g_1} \left(2\mathbf{C}^{(\ell,k+\frac{1}{2})} - \mathbf{C}^{(\ell,k)} \right) - \mathbf{C}^{(\ell,k+\frac{1}{2})} \right)$
 - 14: **end for**
 - 15: **end for**
-

564 A convergence result is next stated, which is inspired from [64] (itself relying on
565 [69, p. 6]), but does not require the differentiability of $g_0 + g_1$.

566 **THEOREM 4.3.** *Let $(\mathbf{C}^{(\ell)})_{\ell \geq 0}$ be a sequence generated by (63). Assume that*
567 *$\text{dom } f \cap \text{dom } g_0 \cap \text{dom } g_1 \neq \emptyset$, $f + g_0 + g_1$ is coercive, and $E = \{\mathbf{C} \in \mathcal{S}_n \mid \mathcal{F}(\mathbf{C}) \leq$*
568 *$\mathcal{F}(\mathbf{C}^{(0)})\}$ is a subset of the relative interior of $\text{dom } g_0 \cap \text{dom } g_1$. Then, the following*
569 *properties hold:*

- 570 (i) $(\mathcal{F}(\mathbf{C}^{(\ell)}))_{\ell \geq 0}$ is a decaying sequence converging to $\widehat{\mathcal{F}} \in \mathbb{R}$.
- 571 (ii) $(\mathbf{C}^{(\ell)})_{\ell \geq 0}$ has a cluster point.
- 572 (iii) Every cluster point $\widehat{\mathbf{C}}$ of $(\mathbf{C}^{(\ell)})_{\ell \geq 0}$ is such that $\mathcal{F}(\widehat{\mathbf{C}}) = \widehat{\mathcal{F}}$ and it is a critical
573 point of \mathcal{F} , i.e. $-\nabla f(\widehat{\mathbf{C}}) - \nabla \mathcal{T}_{\mathbf{S}}(\widehat{\mathbf{C}}) \in \partial(g_0 + g_1)(\widehat{\mathbf{C}})$.

574 *Proof.* First note that $(\mathbf{C}^{(\ell)})_{\ell \geq 0}$ is properly defined by (63) since, for every $\mathbf{C} \in$
575 \mathcal{S}_n^{++} , $\mathcal{G}(\cdot \mid \mathbf{C})$ is a coercive lower-semicontinuous function. It indeed majorizes \mathcal{F}
576 which is coercive, since $f + g_0 + g_1$ has been assumed coercive.

577 (i): As a known property of MM strategies, $(\mathcal{F}(\mathbf{C}^{(\ell)}))_{\ell \geq 0}$ is a decaying sequence [36].
578 Under our assumptions, we have already seen that \mathcal{F} has a minimizer. We deduce
579 that $(\mathcal{F}(\mathbf{C}^{(\ell)}))_{\ell \geq 0}$ is lower bounded, hence convergent.

580 (ii): Since $(\mathcal{F}(\mathbf{C}^{(\ell)}))_{\ell \geq 0}$ is a decaying sequence, $(\forall \ell \geq 0) \mathbf{C}^{(\ell)} \in E$. Since \mathcal{F} is proper,
581 lower-semicontinuous, and coercive, E is a nonempty compact set and $(\mathbf{C}^{(\ell)})_{\ell \geq 0}$ ad-
582 mits a cluster point in E .

583 (iii): If $\widehat{\mathbf{C}}$ is a cluster point of $(\mathbf{C}^{(\ell)})_{\ell \geq 0}$, then there exists a subsequence $(\mathbf{C}^{(\ell_k)})_{k \geq 0}$
584 converging to $\widehat{\mathbf{C}}$. Since E is a nonempty subset of the relative interior of $\text{dom } g_0 \cap$
585 $\text{dom } g_1$ and $g_0 + g_1 \in \Gamma_0(\mathcal{S}_n)$, $g_0 + g_1$ is continuous relative to E [6, Corollary 8.41]. As
586 $f + \mathcal{T}_{\mathbf{S}}$ is continuous on $\text{dom } f \cap \text{dom } \mathcal{T}_{\mathbf{S}} = \mathcal{S}_n^{++}$, \mathcal{F} is continuous relative to E . Hence,
587 $\widehat{\mathcal{F}} = \lim_{k \rightarrow +\infty} \mathcal{F}(\mathbf{C}^{(\ell_k)}) = \mathcal{F}(\widehat{\mathbf{C}})$. On the other hand, by similar arguments applied to

588 sequence $(\mathbf{C}^{(\ell_{k+1})})_{k \geq 0}$, there exists a subsequence $(\mathbf{C}^{(\ell_{k_q+1})})_{q \geq 0}$ converging to some
 589 $\widehat{\mathbf{C}}' \in E$ such that $\widehat{\mathcal{F}} = \mathcal{F}(\widehat{\mathbf{C}}')$. In addition, thanks to (63), we have

590 (64)
$$(\forall \mathbf{C} \in \mathcal{S}_n)(\forall q \in \mathbb{N}) \quad \mathcal{G}(\mathbf{C}^{(\ell_{k_q+1})} \mid \mathbf{C}^{(\ell_{k_q})}) \leq \mathcal{G}(\mathbf{C} \mid \mathbf{C}^{(\ell_{k_q})}).$$

591 By continuity of f and $\nabla \mathcal{T}_{\mathcal{S}}$ on \mathcal{S}_n^{++} and by continuity of $g_0 + g_1$ relative to E ,

592 (65)
$$(\forall \mathbf{C} \in \mathcal{S}_n) \quad \mathcal{G}(\widehat{\mathbf{C}}' \mid \widehat{\mathbf{C}}) \leq \mathcal{G}(\mathbf{C} \mid \widehat{\mathbf{C}}).$$

593 Let us now suppose that $\widehat{\mathbf{C}}$ is not a critical point of \mathcal{F} . Since the subdifferential of
 594 $\mathcal{G}(\cdot \mid \widehat{\mathbf{C}})$ at $\widehat{\mathbf{C}}$ is $\nabla f(\widehat{\mathbf{C}}) + \nabla \mathcal{T}_{\mathcal{S}}(\widehat{\mathbf{C}}) + \partial(g_0 + g_1)(\widehat{\mathbf{C}})$ [6, Corollary 16.48(ii)], the null
 595 matrix does not belong to this subdifferential, which means that $\widehat{\mathbf{C}}$ is not a minimizer
 596 of $\mathcal{G}(\cdot \mid \widehat{\mathbf{C}})$ [6, Theorem 16.3]. It follows from (65) and standard MM properties that
 597 $\mathcal{F}(\widehat{\mathbf{C}}') \leq \mathcal{G}(\widehat{\mathbf{C}}' \mid \widehat{\mathbf{C}}) < \mathcal{G}(\widehat{\mathbf{C}} \mid \widehat{\mathbf{C}}) = \mathcal{F}(\widehat{\mathbf{C}})$. The resulting strict inequality contradicts
 598 the already established fact that $\mathcal{F}(\widehat{\mathbf{C}}') = \mathcal{F}(\widehat{\mathbf{C}})$. \square

599 **5. Numerical Experiments.** This section presents some numerical tests illus-
 600 trating the validity of the proposed algorithms. More specifically, in [Subsection 5.1](#) the
 601 Douglas–Rachford (DR) approach of [Section 3](#) is compared with other state-of-the-
 602 art algorithms previously mentioned, namely Incremental Proximal Descent (IPD)
 603 [54] and ADMM [72], on a problem of covariance matrix estimation. In [Subsec-](#)
 604 [tion 5.2](#), we present an application of the MM approach from [Section 4](#) to a graphical
 605 lasso problem in the presence of noisy data. All the experiments were conducted on
 606 a MacBook Pro equipped with an Intel Core i7 at 2.2 GHz, 16 Gb of RAM (DDR3
 607 1600 MHz), and Matlab R2015b.

608 **5.1. Application to Sparse Covariance Matrix Estimation.** We first con-
 609 sider the application of the DR algorithm from [Section 3](#) to the sparse covariance
 610 matrix estimation problem introduced in [54]. The objective is to retrieve an estimate
 611 of a low rank covariance matrix $\mathbf{Y}^* \in \mathcal{S}_n^+$ from N noisy realizations $(\mathbf{x}^{(i)})_{1 \leq i \leq N}$ of a
 612 Gaussian multivalued random vector with zero mean and covariance matrix $\mathbf{Y}^* + \sigma^2 \mathbf{I}_d$,
 613 with $\sigma > 0$. As we have shown in [Subsection 4.1](#), a solution to this problem can be
 614 obtained by solving the penalized least-squares problem (50), where \mathbf{S} is the empirical
 615 covariance matrix defined in (49), and the regularization terms are $g_0 = \mu_0 \mathcal{R}_1$ and
 616 $g_1 = \mu_1 \|\cdot\|_1$. We propose to compare the performance of the DR approach from [Sub-](#)
 617 [section 3.2](#), with the IPD algorithm [54] and the ADMM procedure [72], for solving
 618 this convex optimization problem.

619 The synthetic data are generated using a procedure similar to the one in [54].
 620 A block-diagonal covariance matrix \mathbf{Y}^* is considered, composed with r blocks with
 621 dimensions $(r_j)_{1 \leq j \leq r}$, so that $n = \sum_{j=1}^r r_j$. The j -th diagonal block of \mathbf{Y}^* reads as
 622 a product $\mathbf{a}_j \mathbf{a}_j^\top$, where the components of $\mathbf{a}_j \in \mathbb{R}^{r_j}$ are randomly drawn on $[-1, 1]$.
 623 The number of observations N is equal to n and $\sigma = 0.1$. The three algorithms
 624 are initialized with $\mathbf{S} + \mathbf{I}_d$, and stopped as soon as a relative decrease criterion on
 625 the objective function is met, i.e. when $|\mathcal{F}_{k+1} - \mathcal{F}_k|/|\mathcal{F}_k| \leq \varepsilon$, $\varepsilon > 0$ being a given
 626 tolerance and \mathcal{F}_k denoting the objective function value at iteration k . The maximum
 627 number of iterations is set to 2000. The penalty parameters μ_1 and μ_0 are chosen
 628 in order to get a reliable estimation of the original covariance matrix. The gradient
 629 stepsize for IPD is set to k^{-1} . In [Algorithm 1](#), α_k is set to 1.5. In ADMM, the initial
 630 Lagrange multiplier is set to a matrix with all entries equal to one, and the parameter
 631 of the proximal step is set to 1.

632 [Figure 1](#) illustrates the quality of the recovered covariance matrices when setting
 633 $\varepsilon = 10^{-10}$. Three different indicators for estimation quality are provided, namely

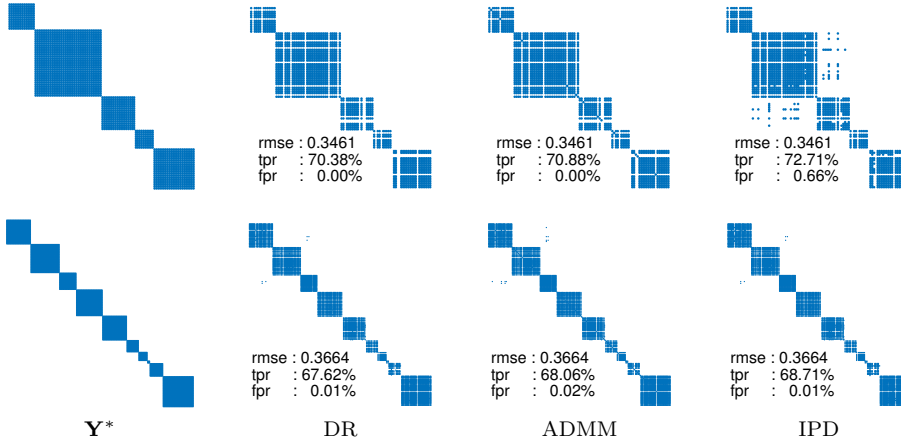


FIG. 1. Original matrix and reconstruction results for DR, ADMM and IPD algorithms, for $n = 100$ (top) and $n = 300$ (bottom).

634 the *true positive rate* (**tpr**), i.e. the correctly recognized non-zero entries, the *false*
635 *positive rate* (**fpr**), i.e. the entries erroneously added to the support of the matrix,
636 and the *relative mean square error* (**rmse**), computed as $\|\mathbf{Y}_{\text{rec}} - \mathbf{Y}^*\|_{\text{F}}^2 / \|\mathbf{Y}^*\|_{\text{F}}^2$, with
637 \mathbf{Y}_{rec} the recovered matrix. Note that the two first measurements are employed when
638 the main interest lies in the recovery of the matrix support. A visual inspection shows
639 that the three methods provide similar results in terms of matrix support estimation.
640 Moreover, the reconstruction error as well as the values of **fpr** and **tpr** slightly differ.

TABLE 4

Comparison in terms of convergence speed between DR, ADMM and IPD procedures. The enlighten times refers to the shortest ones.

	$n = 100, \mu_0 = 0.2, \mu_1 = 0.1, r = 5$ $\{r_j\} = \{14, 36, 18, 10, 22\}$			$n = 300, \mu_0 = 0.01, \mu_1 = 0.12$ $r = 10, \{r_j\} = \{39, 46, 27, 42, 39, 19, 14, 4, 21, 49\}$		
	DR	ADMM	IPD	DR	ADMM	IPD
ε	Time(iter)	Time(iter)	Time(iter)	Time(iter)	Time(iter)	Time(iter)
10^{-6}	0.03 (23)	0.02 (17)	0.18 (167)	0.14 (17)	0.11 (14)	1.34 (170)
10^{-7}	0.03 (27)	0.02 (21)	0.58 (533)	0.32 (38)	0.34 (42)	4.35 (548)
10^{-8}	0.03 (30)	0.04 (34)	1.83 (685)	0.81 (95)	0.91 (115)	13.72 (1748)
10^{-9}	0.06 (56)	0.06 (54)	2.16 (2000)	1.79 (211)	2.06 (258)	15.70 (2000)
10^{-10}	0.07 (59)	0.07 (58)	2.16 (2000)	5.23 (620)	5.45 (686)	15.68 (2000)

641 **Table 4** presents the comparative performance of the algorithms in terms of com-
642 putation time (in second) and iteration number (averaged on 20 noise realizations),
643 for two scenarios corresponding to distinct problem sizes and block distributions. It
644 can be observed that the behaviors of ADMM and DR are similar, while IPD requires
645 more iterations and time to reach the same precision. Furthermore, the latter fails
646 to reach a high precision in the allowed maximum number of iterations, for both
647 examples.

648 **5.2. Application to Robust Graphical Lasso.** Let us now illustrate the
649 applicability of the MM approach presented in [Subsection 4.3](#) to the problem of

650 precision matrix estimation introduced in (51). The test datasets have been gener-
 651 ated by using the code available at [http://stanford.edu/boyd/papers/admm/covsel/](http://stanford.edu/boyd/papers/admm/covsel/covsel_example.html)
 652 [covsel_example.html](http://stanford.edu/boyd/papers/admm/covsel/covsel_example.html). A sparse precision matrix \mathbf{C}^* of dimension $n \times n$ is randomly
 653 created, where the number of non-zero entries is chosen as a proportion $p \in]0, 1[$ of
 654 the total number n^2 . Then, N realizations $(\mathbf{x}^{(i)})_{1 \leq i \leq N}$ of a Gaussian multivalued
 655 random variable with zero mean and covariance $\mathbf{Y}^* = (\mathbf{C}^*)^{-1}$ are generated. Gaus-
 656 sian noise with zero mean and covariance $\sigma^2 \mathbf{I}_d$, $\sigma > 0$, is finally added to the $\mathbf{x}^{(i)}$'s,
 657 so that the covariance matrix Σ associated with the input data reads as in (47) with
 658 $\mathbf{A} = \mathbf{I}_d$. As explained in Subsection 4.1, the estimation of \mathbf{C}^* can be performed by
 659 using the MM algorithm from Subsection 4.3 based on the minimization of the non-
 660 convex cost (51) with regularization functions $g_1 = \mu_1 \|\cdot\|_1$, $\mu_1 > 0$, and $(\forall \mathbf{C} \in \mathcal{S}_n^{++})$
 661 $g_0(\mathbf{C}) = \mu_0 \mathcal{R}_1(\mathbf{C}^{-1})$, $\mu_0 > 0$. The computation of $\text{prox}_{\gamma(\varphi+\psi)}$ with $\gamma \in]0, +\infty[$ re-
 662 lated to this particular choice for g_0 and function φ given by (57) and (55) leads to
 663 the search of the only positive root of a polynomial of degree 4.

664 A synthetic dataset of size $n = 100$ is created, where matrix \mathbf{C}^* has 20 off-
 665 diagonal non-zero entries (i.e., $p = 10^{-3}$) and the corresponding covariance matrix
 666 has condition number 0.125. $N = 1000$ realizations are used to compute the empirical
 667 covariance matrix \mathbf{S} . In our MM algorithm, the inner stopping criterion (line 7 in
 668 Algorithm 2) is based on the relative difference of majorant function values with a
 669 tolerance of 10^{-10} , while the outer cycle is stopped when the relative difference of
 670 the objective function values falls below 10^{-8} . The DR algorithm is used to solve the
 671 inner subproblems, by using parameters $(\forall \ell) \gamma_\ell = 1$, $(\forall k) \alpha_{\ell,k} = 1$ (see Algorithm 2,
 672 lines 4–13). The allowed maximum inner (resp. outer) iteration number is 2000 (resp.
 673 20). The quality of the results is quantified in terms of **fpr** on the precision matrix and
 674 **rmse** with respect to the true covariance matrix. The parameters μ_1 and μ_0 are set in
 675 order to obtain the best reconstruction in terms of **rmse**. For eight values of the noise
 676 standard deviation σ , Figure 2 illustrates the reconstruction quality (averaged on 20
 677 noise realizations) obtained with our method, as well as two other approaches that
 678 do not take into account the noise in their formulation, namely the classical GLASSO
 679 approach from [12], which amounts to solve (1) with $f = -\log \det$, $g = \mu_1 \|\cdot\|_1$,
 680 and the DR approach described in Section 3, in the formulation given by (1) with
 681 $f = -\log \det$, $(\forall \mathbf{C} \in \mathcal{S}_n^{++}) g(\mathbf{C}) = \mu_0 \mathcal{R}_1(\mathbf{C}^{-1}) + \mu_1 \|\mathbf{C}\|_1$. For the DR approach,
 682 $\text{prox}_{\gamma(\varphi+\psi)}$ with $\gamma \in]0, +\infty[$ is given by the fourth line of Table 2 (when $p = 1$).

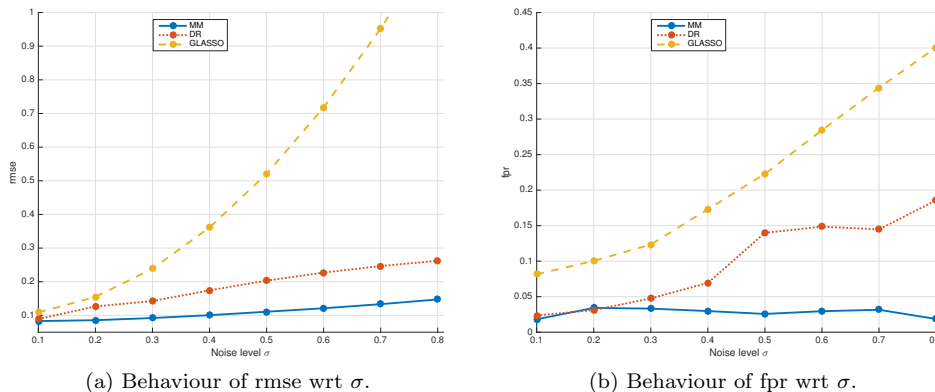


FIG. 2. Estimation results for different noise levels in terms of *rmse* (left) and *fpr* (right) for MM, GLASSO and DR approaches.

683 As expected, as the noise variance increases the reconstruction quality deteriorates. The GLASSO procedure is strongly impacted by the presence of noise, whereas
 684 the MM approach achieves better results, also when compared with DR algorithm.
 685 Moreover, the MM algorithm significantly outperforms both other methods in terms
 686 of support reconstruction, revealing itself very robust with respect to an increasing
 687 level of noise.
 688

689 **6. Conclusions.** In this work, various proximal tools have been introduced to
 690 deal with optimization problems involving real symmetric matrices. We have focused
 691 on the variational framework (1) which is closely related to the computation of a
 692 proximity operator with respect to a Bregman divergence. It has been assumed that
 693 f in (3) is a convex spectral function, and g reads as $g_0 + g_1$, where g_0 is a spectral
 694 function. We have given a fully spectral solution in Section 2 when $g_1 \equiv 0$, and,
 695 in particular, Corollary 2.6 could be useful for developing algorithms involving prox-
 696 imity operators in other metrics than the Frobenius one. When $g_1 \neq 0$, a proximal
 697 iterative approach has been presented, which is grounded on the use of the Douglas–
 698 Rachford procedure. As illustrated by the tables of proximity operators provided
 699 for a wide range of choices for f and g_0 , the main advantage of the proposed algo-
 700 rithm is its great flexibility. The proposed framework also has allowed us to propose
 701 a nonconvex formulation of the precision matrix estimation problem arising in the
 702 context of noisy graphical lasso. The nonconvexity of the obtained objective function
 703 has been circumvented through a Majorization–Minimization approach, each step
 704 of which consists of solving a convex problem by a Douglas-Rachford sub-iteration.
 705 Comparisons with state-of-the-art solutions have demonstrated the robustness of the
 706 proposed method. It is worth mentioning that all the results presented in this paper
 707 can be easily extended to complex Hermitian matrices.

708

REFERENCES

- 709 [1] F. J. ARAGÓN ARTACHO AND J. M. BORWEIN, *Global convergence of a non-convex Douglas–*
 710 *Rachford iteration*, J. Global Optim., 57 (2013), pp. 753–769, [https://doi.org/10.1007/](https://doi.org/10.1007/s10898-012-9958-4)
 711 [s10898-012-9958-4](https://doi.org/10.1007/s10898-012-9958-4).
 712 [2] M. S. ASLAN, X.-W. CHEN, AND H. CHENG, *Analyzing and learning sparse and scale-free*
 713 *networks using Gaussian graphical models*, J. Mach. Learn. Res., 1 (2016), pp. 99–109,
 714 <https://doi.org/10.1007/s41060-016-0009-y>.
 715 [3] O. BANERJEE, L. EL GHAOUI, AND A. D’ASPREMONT, *Model selection through sparse maximum*
 716 *likelihood estimation for multivariate Gaussian or binary data*, J. Mach. Learn. Res., 9
 717 (2008), pp. 485–516.
 718 [4] H. H. BAUSCHKE, J. M. BORWEIN, AND P. L. COMBETTES, *Essential smoothness, essential strict*
 719 *convexity, and Legendre functions in Banach spaces*, Comm. Contemp. Math, 3 (2001),
 720 pp. 615–647.
 721 [5] H. H. BAUSCHKE, J. M. BORWEIN, AND P. L. COMBETTES, *Bregman monotone optimization*
 722 *algorithms*, SIAM J. Control Optim., 42 (2003), pp. 596–636, [https://doi.org/10.1137/](https://doi.org/10.1137/S0363012902407120)
 723 [S0363012902407120](https://doi.org/10.1137/S0363012902407120).
 724 [6] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in*
 725 *Hilbert Spaces*, Springer International Publishing, 2nd ed., 2017, [https://doi.org/10.1007/](https://doi.org/10.1007/978-3-319-48311-5)
 726 [978-3-319-48311-5](https://doi.org/10.1007/978-3-319-48311-5).
 727 [7] H. H. BAUSCHKE, P. L. COMBETTES, AND D. NOLL, *Joint minimization with alternating Breg-*
 728 *man proximity operators*, Pac. J. Optim., 2 (2006), pp. 401–424.
 729 [8] A. BENFENATI AND V. RUGGIERO, *Inexact Bregman iteration with an application to Poisson*
 730 *data reconstruction*, Inverse Problems, 29 (2013), pp. 1–32.
 731 [9] A. BENFENATI AND V. RUGGIERO, *Inexact Bregman iteration for deconvolution of superimposed*
 732 *extended and point sources*, Commun. Nonlinear Sci. Numer. Simul., 20 (2015), pp. 882 –
 733 896, <https://doi.org/http://dx.doi.org/10.1016/j.cnsns.2014.06.045>.
 734 [10] I. BENGTSOON AND K. ZYCZKOWSKI, *Geometry of Quantum States: An Introduction to Quan-*
 735 *tum Entanglement*, Cambridge University Press, Cambridge, 002 2006, [https://doi.org/10.](https://doi.org/10.1017/9780521876223)

- 736 [1017/CBO9780511535048](https://doi.org/10.1017/CBO9780511535048).
- 737 [11] J. BORWEIN AND A. LEWIS, *Convex Analysis and Nonlinear Optimization*, Springer, 2014.
- 738 [12] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and*
739 *statistical learning via the alternating direction method of multipliers*, Found. Trends Mach.
740 Learn., 3 (2011), pp. 1–122, <https://doi.org/10.1561/22000000016>.
- 741 [13] L. M. BREGMAN, *The Relaxation Method of Finding the Common Point of Convex Sets and Its*
742 *Application to the Solution of Problems in Convex Programming*, USSR Computational
743 Mathematics and Mathematical Physics, 7 (1967), pp. 200–217.
- 744 [14] C. BRUNE, A. SAWATZKY, AND M. BURGER, *Primal and dual Bregman methods with application*
745 *to optical nanoscopy*, Int. J. Comput. Vis., 92 (2011), pp. 211–229, [https://doi.org/10.](https://doi.org/10.1007/s11263-010-0339-5)
746 [1007/s11263-010-0339-5](https://doi.org/10.1007/s11263-010-0339-5).
- 747 [15] M. BURGER, A. SAWATZKY, AND G. STEIDL, *First Order Algorithms in Variational Image*
748 *Processing*, Springer International Publishing, Cham, 2016, pp. 345–407, [https://doi.org/](https://doi.org/10.1007/978-3-319-41589-5_10)
749 [10.1007/978-3-319-41589-5_10](https://doi.org/10.1007/978-3-319-41589-5_10).
- 750 [16] J.-F. CAI, E. J. CANDS, AND Z. SHEN, *A singular value thresholding algorithm for matrix com-*
751 *pletion*, SIAM J. Optim., 20 (2010), pp. 1956–1982, <https://doi.org/10.1137/080738970>.
- 752 [17] T. CAI, W. LIU, AND X. LUO, *A constrained ℓ_1 minimization approach to sparse precision*
753 *matrix estimation*, J. Am. Stat. Assoc., 106 (2011), pp. 594–607, [https://doi.org/10.1198/](https://doi.org/10.1198/jasa.2011.tm10155)
754 [jasa.2011.tm10155](https://doi.org/10.1198/jasa.2011.tm10155).
- 755 [18] V. CHANDRASEKARAN, P. A. PARRILO, AND A. S. WILLSKY, *Latent variable graphical model*
756 *selection via convex optimization*, Ann. Statist., 40 (2012), pp. 1935–1967, [https://doi.org/](https://doi.org/10.1214/11-AOS949)
757 [10.1214/11-AOS949](https://doi.org/10.1214/11-AOS949).
- 758 [19] R. CHARTRAND, *Nonconvex splitting for regularized low-rank + sparse decomposition*, IEEE
759 Trans. Signal Process., 60 (2012), pp. 5810–5819.
- 760 [20] C. CHAUX, P. L. COMBETTES, J.-C. PESQUET, AND V. R. WAJS, *A variational formulation for*
761 *frame-based inverse problems*, Inverse Problems, 23 (2007), p. 1495.
- 762 [21] C. CHAUX, J.-C. PESQUET, AND N. PUSTELNIK, *Nested iterative algorithms for convex con-*
763 *strained image recovery problem*, SIAM J. Imaging Sci., 2 (2009), pp. 730–762.
- 764 [22] E. CHOUZENOUX AND J.-C. PESQUET, *Convergence Rate Analysis of the Majorize-Minimize*
765 *Subspace Algorithm*, IEEE Signal Process. Lett., 23 (2016), pp. 1284 – 1288, [https://doi.](https://doi.org/10.1109/LSP.2016.2593589)
766 [org/10.1109/LSP.2016.2593589](https://doi.org/10.1109/LSP.2016.2593589).
- 767 [23] P. L. COMBETTES AND J.-C. PESQUET, *A Douglas-Rachford splitting approach to nonsmooth*
768 *convex variational signal recovery*, IEEE J. Sel. Topics Signal Process., 1 (2007), pp. 564–
769 574.
- 770 [24] P. L. COMBETTES AND J.-C. PESQUET, *Proximal Splitting Methods in Signal Processing*, in
771 Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, 2011,
772 pp. 185–212, <https://doi.org/10.1007/978-1-4419-9569-8>.
- 773 [25] L. CONDAT, *Fast projection onto the simplex and the ℓ_1 ball*, Math. Program., 158 (2016),
774 pp. 575–585, <https://doi.org/10.1007/s10107-015-0946-6>.
- 775 [26] R. M. CORLESS, G. H. GONNET, D. E. G. HARE, D. J. JEFFREY, AND D. E. KNUTH, *On*
776 *the Lambert W function*, Adv. Comput. Math., 5 (1996), pp. 329–359, [https://doi.org/10.](https://doi.org/10.1007/BF02124750)
777 [1007/BF02124750](https://doi.org/10.1007/BF02124750).
- 778 [27] T. COVER AND J. THOMAS, *Elements of Information Theory*, A Wiley-Interscience publication,
779 Wiley, 2006.
- 780 [28] A. D’ASPROMONT, O. BANERJEE, AND L. E. GHAOUI, *First-order methods for sparse covariance*
781 *selection*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 56–66, [https://doi.org/10.1137/](https://doi.org/10.1137/060670985)
782 [060670985](https://doi.org/10.1137/060670985).
- 783 [29] A. DEMPSTER, *Covariance selection*, Biometrics, 28 (1972), pp. 157–175.
- 784 [30] J. C. DUCHI, S. GOULD, AND D. KOLLER, *Projected Subgradient Methods for Learning Sparse*
785 *Gaussians*, in UAI 2008, Proceedings of the 24th Conference in Uncertainty in Artificial
786 Intelligence, Helsinki, Finland, July 9–12, 2008, 2008, pp. 145–152.
- 787 [31] J. FRIEDMAN, T. HASTIE, AND R. TIBSHIRANI, *Sparse inverse covariance estimation with the*
788 *graphical lasso*, Biostatistics, 9 (2008), pp. 432–441, [https://doi.org/10.1093/biostatistics/](https://doi.org/10.1093/biostatistics/kxm045)
789 [kxm045](https://doi.org/10.1093/biostatistics/kxm045).
- 790 [32] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for l_1 -regularized problems*, SIAM J.
791 Imaging Sci., 2 (2009), pp. 323–343, <https://doi.org/10.1137/080725891>.
- 792 [33] J. GUO, E. LEVINA, G. MICHAILIDIS, AND J. ZHU, *Joint estimation of multiple graphical models*,
793 Biometrika, 98 (2011), p. 1, <https://doi.org/10.1093/biomet/asq060>.
- 794 [34] G. HARDY, J. LITTLEWOOD, AND G. PÓLYA, *Inequalities*, Cambridge Mathematical Library,
795 Cambridge University Press, 1952.
- 796 [35] D. R. HUNTER AND K. LANGE, *A tutorial on MM algorithms*, Amer. Statist., 58 (2004), pp. 30–
797 37, <https://doi.org/10.1198/0003130042836>.

- 798 [36] M. W. JACOBSON AND J. A. FESSLER, *An expanded theoretical treatment of iteration-dependent*
799 *majorize-minimize algorithms*, IEEE Trans. Image Process., 16 (2007), pp. 2411–2422,
800 <https://doi.org/10.1109/TIP.2007.904387>.
- 801 [37] N. KOMODAKIS AND J. C. PESQUET, *Playing with duality: An overview of recent primal–dual*
802 *approaches for solving large-scale optimization problems*, IEEE Signal Process. Mag., 32
803 (2015), pp. 31–54, <https://doi.org/10.1109/MSP.2014.2377273>.
- 804 [38] A. S. LEWIS, *Convex analysis on the Hermitian matrices*, SIAM J. Optim., 6 (1996), pp. 164–
805 177, <https://doi.org/10.1137/0806009>.
- 806 [39] G. LI AND T. K. PONG, *Douglas–Rachford splitting for nonconvex optimization with application*
807 *to nonconvex feasibility problems*, Math. Programm., 159 (2016), pp. 371–401, <https://doi.org/10.1007/s10107-015-0963-5>.
- 808 [40] L. LI AND K.-C. TOH, *An inexact interior point method for ℓ_1 -regularized sparse covari-*
809 *ance selection*, Math. Program. Comput., 2 (2010), pp. 291–315, <https://doi.org/10.1007/s12532-010-0020-6>.
- 812 [41] P. L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*,
813 SIAM J. Numer. Anal., 16 (1979), pp. 964–979, <https://doi.org/10.1137/0716071>.
- 814 [42] Z. LU, *Smooth optimization approach for sparse covariance selection*, SIAM J. Optim., 19
815 (2009), pp. 1807–1827, <https://doi.org/10.1137/070695915>.
- 816 [43] Z. LU, *Adaptive first-order methods for general sparse inverse covariance selection*, SIAM J.
817 Matrix Anal. Appl., 31 (2010), pp. 2000–2016, <https://doi.org/10.1137/080742531>.
- 818 [44] S. MA, L. XUE, AND H. ZOU, *Alternating direction methods for latent variable Gaussian graph-*
819 *ical model selection*, Neural Comput., 25 (2013), pp. 2172–2198, https://doi.org/10.1162/NECO_a.00379.
- 820 [45] J. R. MAGNUS AND H. NEUDECKER, *Matrix Differential Calculus with Applications in Statistics*
821 *and Econometrics*, John Wiley, second ed., 1999.
- 823 [46] A. W. MARSHALL, I. OLKIN, AND B. C. ARNOLD, *Inequalities: Theory of Majorization*
824 *and its Applications*, vol. 143, Springer, second ed., 2011, <https://doi.org/10.1007/978-0-387-68276-1>.
- 826 [47] R. MAZUMDER AND T. HASTIE, *The graphical lasso: New insights and alternatives*, Electron.
827 J. Stat., 6 (2012), pp. 2125–2149, <https://doi.org/10.1214/12-EJS740>.
- 828 [48] N. MEINSHAUSEN AND P. BHLMANN, *High-dimensional graphs and variable selection*
829 *with the lasso*, Ann. Statist., 34 (2006), pp. 1436–1462, <https://doi.org/10.1214/009053606000000281>.
- 830 [49] J. MOREAU, *Proximit et dualit dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965),
831 pp. 273–299.
- 833 [50] Y. NESTEROV, *Smooth minimization of non-smooth functions*, Math. Programm., 103 (2005),
834 pp. 127–152, <https://doi.org/10.1007/s10107-004-0552-5>.
- 835 [51] N. PARIKH AND S. BOYD, *Proximal algorithms*, Found. Trends Optim., 1 (2014), pp. 127–239,
836 <https://doi.org/10.1561/2400000003>.
- 837 [52] J.-C. PESQUET AND N. PUSTELNIK, *A parallel inertial proximal optimization method*, Pac. J.
838 Optim., 8 (2012), pp. 273–305.
- 839 [53] P. RAVIKUMAR, M. J. WAINWRIGHT, G. RASKUTTI, AND B. YU, *High-dimensional covariance*
840 *estimation by minimizing ℓ_1 -penalized log-determinant divergence*, Electron. J. Statist., 5
841 (2011), pp. 935–980, <https://doi.org/10.1214/11-EJS631>.
- 842 [54] E. RICHARD, P. ANDRE SAVALLE, AND N. VAYATIS, *Estimation of simultaneously sparse and low*
843 *rank matrices*, in Proceedings of the 29th International Conference on Machine Learning
844 (ICML-12), ACM, 2012, pp. 1351–1358.
- 845 [55] R. ROCKAFELLAR, *Convex Analysis*, Princeton landmarks in mathematics and physics, Prince-
846 ton University Press, 1970.
- 847 [56] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer-Verlag, 1st ed., 1997.
- 848 [57] A. J. ROTHMAN, P. J. BICKEL, E. LEVINA, AND J. ZHU, *Sparse permutation invariant co-*
849 *variance estimation*, Electron. J. Statist., 2 (2008), pp. 494–515, <https://doi.org/10.1214/08-EJS176>.
- 850 [58] K. SCHEINBERG, S. MA, AND D. GOLDFARB, *Sparse inverse covariance selection via alternating*
851 *linearization methods*, in Advances in Neural Information Processing Systems 23, 2010,
852 pp. 2101–2109.
- 854 [59] Y. SUN, P. BABU, AND D. P. PALOMAR, *Majorization-Minimization algorithms in signal pro-*
855 *cessing, communications, and machine learning*, IEEE Trans. Signal Process., 65 (2017),
856 pp. 794–816, <https://doi.org/10.1109/TSP.2016.2601299>.
- 857 [60] M. E. TIPPING, *Sparse Bayesian learning and the relevance vector machine*, J. Mach. Learn.
858 Res., 1 (2001), pp. 211–244, <https://doi.org/10.1162/15324430152748236>.
- 859 [61] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Probing the Pareto frontier for basis pursuit solu-*

- 860 *tions*, SIAM J. Sci. Comput., 31 (2009), pp. 890–912, <https://doi.org/10.1137/080714488>.
- 861 [62] C. WANG, D. SUN, AND K.-C. TOH, *Solving log-determinant optimization problems by a*
862 *Newton-CG primal proximal point algorithm*, SIAM J. Optim., 20 (2010), pp. 2994–3013,
863 <https://doi.org/10.1137/090772514>.
- 864 [63] D. P. WIPF AND B. D. RAO, *Sparse Bayesian learning for basis selection*, IEEE Trans. Signal
865 Process., 52 (2004), pp. 2153–2164, <https://doi.org/10.1109/TSP.2004.831016>.
- 866 [64] C. F. J. WU, *On the convergence properties of the EM algorithm*, Ann. Statist., 11 (1983),
867 pp. 95–103, <https://doi.org/10.1214/aos/1176346060>.
- 868 [65] S. YANG, Z. LU, X. SHEN, P. WONKA, AND J. YE, *Fused multiple graphical lasso*, SIAM J.
869 Optim., 25 (2015), pp. 916–943, <https://doi.org/10.1137/130936397>.
- 870 [66] W. YIN, S. OSHER, D. GOLDFARB, AND J. DARBON, *Bregman iterative algorithms for ℓ_1 -*
871 *minimization with applications to compressed sensing*, SIAM J. Imaging Sci., 1 (2008),
872 pp. 143–168, <https://doi.org/10.1137/070703983>.
- 873 [67] M. YUAN AND Y. LIN, *Model selection and estimation in the Gaussian graphical model*,
874 Biometrika, 94 (2007), p. 19, <https://doi.org/10.1093/biomet/asm018>.
- 875 [68] X. YUAN, *Alternating direction methods for sparse covariance selection*, (2009), [http://www.](http://www.optimization-online.org/DBFILE/2009/09/2390.pdf)
876 [optimization-online.org/DBFILE/2009/09/2390.pdf](http://www.optimization-online.org/DBFILE/2009/09/2390.pdf).
- 877 [69] W. I. ZANGWILL, *Nonlinear programming : a unified approach*, Englewood Cliffs, N.J. :
878 Prentice-Hall, 1969.
- 879 [70] X. ZHANG, M. BURGER, X. BRESSON, AND S. OSHER, *Bregmanized nonlocal regularization for*
880 *deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276,
881 <https://doi.org/10.1137/090746379>.
- 882 [71] X. ZHANG, M. BURGER, AND S. OSHER, *A unified primal-dual algorithm framework based*
883 *on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46, [https://doi.org/10.1007/](https://doi.org/10.1007/s10915-010-9408-8)
884 [s10915-010-9408-8](https://doi.org/10.1007/s10915-010-9408-8).
- 885 [72] S. ZHOU, N. XIU, Z. LUO, AND L. KONG, *Sparse and low-rank covariance matrices estimation*,
886 (2014), <https://arxiv.org/abs/1407.4596>.