IMA Journal of Numerical Analysis (2013) Page 1 of 16
doi:10.1093/imanum/drt056

Correlation matrix nearness and completion under observation uncertainty

Carlos M. Alaíz
Departamento de Ingeniería Informática, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain
carlos.alaiz@uam.es

and

Francesco Dinuzzo and Suvrit Sra*
Max Planck Institute for Intelligent Systems, Spemannstrasse 38, 72076 Tübingen, Germany
*Corresponding author: suvrit@tuebingen.mpg.de fd@is.mpg.de

[Received on 18 March 2013; revised on 25 October 2013]

This paper introduces the paradigm of optimization under uncertainty for modelling and solving matrix nearness problems. In particular, it considers the concrete problem of recovering correlation matrices from uncertain observations by introducing two different approaches to tackling uncertainty. The first approach invokes the framework of robust optimization to construct low error solutions that are immune to worst-case uncertainty in the input. The second approach takes a less pessimistic view on uncertainty, and considers a situation where instead of the worst one, it suffices to use *any* matrix in the uncertainty set. We formulate both our approaches as convex (possibly nonsmooth) optimization problems. Thereafter, we show how to exploit problem structure to obtain efficient iterative first-order algorithms. We present several numerical results on both nearness and completion versions of the optimization problem; our results highlight the effectiveness of our proposed algorithmic techniques.

Keywords: matrix nearness; matrix completion; robust optimization; correlation matrices.

1. Introduction

Real applications never have access to perfect data: noise, imprecision and incompleteness invariably complicate matters. Since uncertainty is unavoidable, it is important to design models and algorithms that account for it, preferably without being too sensitive to minor variations in data. This pragmatic viewpoint rings particularly true for matrix nearness and completion problems because they involve a noisy, possibly incomplete matrix of uncertain observations that must be processed to fulfil desired properties such as symmetry, semidefiniteness, etc.—see, e.g., the eminent survey by Higham (1989).

A classical way to model uncertainty is via probability theory, as is notably done in stochastic programming (Dantzig, 1955; Shapiro *et al.*, 2009), which deals with uncertainty by modelling input data as random variables. Although attractive and elegant, such probabilistic approaches can be impractical: precise knowledge of probability distributions is rare, and even when we have access to such distributions, the associated mathematical models may be computationally prohibitive. This viewpoint is developed more deeply by Bertsimas & Thiele (2006), who advocate *Robust Optimization* (Bertsimas & Thiele, 2006; Ben-Tal *et al.*, 2009) as an alternative approach to cope with uncertainty. Indeed, robust optimization offers a powerful framework that accounts for observation uncertainty and strives to obtains solutions that are 'worst-case' optimal.

In light of the above background, our paper makes four key contributions: (i) it introduces robust optimization to the field of matrix nearness; (ii) it illustrates the robust paradigm by applying it to

the concrete problem of *correlation matrix nearness*; (iii) it goes beyond robust optimization and offers practitioners an alternative model of uncertainty that is based on an 'explorative' rather than a worst-case view; and (iv) it formulates the resulting matrix nearness problems as convex optimization problems, for which it subsequently develops efficient iterative algorithms.

As our presentation readily suggests, tackling uncertainty via robust or explorative approaches applies to any matrix nearness problem. But a concrete example will illustrate our ideas better; we use the 'nearest correlation matrix' (NCM) problem (Higham, 2002; Gentle, 2007, Section 9.4.6) as our example problem. NCM is not only important in many applications (see, e.g., Bhansali & Wise, 2001; Higham, 2002), but it also admits fairly nontrivial yet tractable uncertainty models.

Note added in revision. This paper is one of the first to formally introduce robust optimization for matrix nearness problems. We thank a referee for bringing to our attention the preprint (Li *et al.*, 2013), which studies robust least-squares SDP problems, and seems to have been written around the same time (in 2012) when we were preparing our paper. Li *et al.* (2013) study robust SDP projection problems whose constraints may have uncertainty; we consider robust SDP projection problems where the point being projected has uncertain entries. In addition, to standard worst-case uncertainty, we also introduce and study an alternative 'explorative' notion.

Notation. Before we begin, let us fix basic notation. Without loss of generality, all matrices considered are real. The operator $[x]_+ = \max\{x, 0\}$ denotes the non-negative part of x. We denote by $\mathscr{S}^n$ the set of $n \times n$ real symmetric matrices, and by $\mathscr{S}^n_+ := \{X \in \mathscr{S}^n \mid X \succeq 0\}$ the convex cone of symmetric positive semidefinite matrices. Let $S^n_D := \{X \in \mathscr{S}^n \mid x_{ii} = 1, i = 1, \ldots n\}$ be the set of matrices with unit diagonal. Using this notation, the set of correlation matrices is $\mathscr{C}_n := \mathscr{S}^n_+ \cap S^n_D$. We also use the Frobenius norm $\|X\|_F = \sqrt{\sum_{ij} x^2_{ij}}$, the operator 2-norm $\|X\|_2 = \sigma_{\max}(X)$ (largest singular value) and the vector ℓ_1-norm $\|X\|_1 = \sum_{ij} |x_{ij}|$. Finally, for matrices A and B of the same dimension, we use $A \circ B$ to denote their Schur (elementwise) product.

2. Background: observation uncertainty

The basic NCM problem of Higham (2002) starts with an $n \times n$ real symmetric matrix C, which may fail to be a correlation matrix (i.e., a symmetric positive semidefinite matrix with ones on the diagonal) due to measurement errors, noise or otherwise. To fix this deficiency, NCM seeks the 'nearest' matrix in the set $\mathscr{C}_n$. Higham (2002) casts NCM as the following orthogonal projection task:

$$\text{minimize} \quad \|X - C\|^2_F, \quad \text{s.t. } X \in \mathscr{C}_n. \tag{2.1}$$

This formulation, however, has a limitation: it requires the observation matrix C to be available *exactly*. But C contains measurements that might have different levels of uncertainty; in fact, some of its entries might be even missing. One way to deal with such uncertain C is to consider a W-weighted version of NCM (Higham, 2002)

$$\text{minimize} \quad \|W \circ (X - C)\|^2_F, \tag{2.2}$$

where W is a matrix of non-negative weights that expresses our confidence in the measurements. From a practical standpoint, however, it is not always clear how to set the weights W in a principled manner, because obtaining these weights may itself be a difficult estimation problem.

The concern here is that the observation matrix C might be an arbitrary sample from some *uncertainty set* $\mathscr{U}$. It seems therefore improvident to find *the* NCM to an arbitrary C that we happened to pick. What we rather want is a solution that works well for the entire uncertainty set $\mathscr{U}$. The robust

optimization paradigm (Bertsimas & Thiele, 2006; Ben-Tal *et al.*, 2009) translates this wish into a problem designed to protect against the 'worst-case', namely,

$$\min_{X \in \mathscr{C}_n} \max_{C \in \mathscr{U}} \quad \|X - C\|_F^2. \tag{2.3}$$

If we had the luxury of exact observations, then the uncertainty set $\mathscr{U}$ would be the singleton $\{C\}$, whereby the inner maximization in (2.3) would disappear, yielding the original NCM problem (2.1). More realistically, $\mathscr{U} \neq \{C\}$, and in this case, unless $\mathscr{U}$ has a convenient structure, the inner maximization in (2.3) may be intractable. Thus, we must balance faithful modelling of uncertainty with computational practicality. We introduce one such practical choice of $\mathscr{U}$ in Section 3.

Complementary to the robust paradigm, we introduce another model for incorporating uncertainty: instead of a worst-case optimal solution, we seek a 'best-case' solution by solving the problem

$$\min_{X \in \mathscr{C}_n} \min_{C \in \mathscr{U}} \quad \|X - C\|_F^2. \tag{2.4}$$

Problem (2.4) uses what we call an *explorative uncertainty* model. The key idea behind this model is that sometimes we might not know exactly which uncertainty set to use. We might, therefore, start with a crude uncertainty set $\mathscr{U}$, obtain the best-case solution under it, and perhaps use this solution to refine our uncertainty set. Thus, we call this approach to handling uncertainty an 'explorative' approach. Such exploration is also helpful if we expect the (uncertain) observation matrix to lie in the uncertainty set.

3. Problem formulation

We are now ready to present a more precise formulation of the uncertainty models for both the robust (Section 3.1) and explorative (Section 3.2) versions of the NCM problem.

3.1 *Robust correlation matrix nearness*

A convenient and easy to interpret model of uncertainty is a collection of intervals. More precisely, we assume that the observation matrix C lies in the uncertainty set given by the box

$$\mathscr{U} = \mathscr{B} := \{Y \in \mathscr{S}^n \mid L \leqslant Y \leqslant U\}, \tag{3.1}$$

where $L, U \in \mathscr{S}^n$ specify lower and upper bounds on entries of C (inequalities are to be understood elementwise). Under this interval-based uncertainty model, problem (2.3) becomes

$$\min_{X \in \mathscr{C}_n} \max_{C \in \mathscr{B}} \quad \|X - C\|_F^2, \tag{3.2}$$

and, henceforth, we refer to (3.2) as the *Robust Nearest Correlation Matrix* (R-NCM) problem.

REMARK 3.1 Other possible uncertainty sets may include the spectral ball

$$\mathscr{U} = \{Y \in \mathscr{S}^n \mid \|Y\|_2 \leqslant \gamma\} \quad \text{for some } \gamma > 0.$$

More generally, one may consider uncertainties in the spectrum of the observation matrix, although such models might be somehow less interpretable than the box.

Albeit convex, (3.2) is not in a form conducive to efficient optimization. Lemma 3.2 shows how to transform it into an equivalent, but more convenient convex optimization problem.

LEMMA 3.2 (R-NCM–convex formulation) Let $\mathscr{B}$ be as given by (3.1). Let M and D be matrices that denote the midpoints and radii of the uncertainty intervals, respectively, that is

$$M := [m_{ij}]_{i,j=1}^{n} := \left[\frac{l_{ij}+u_{ij}}{2}\right]_{i,j=1}^{n}, \quad D := [d_{ij}]_{i,j=1}^{n} := \left[\frac{u_{ij}-l_{ij}}{2}\right]_{i,j=1}^{n}.$$

Then, (3.2) has a unique optimal solution X^*, obtainable by solving the convex optimization problem

$$\min_{X \in \mathscr{C}_n} \quad \|X - M\|_F^2 + 2\|D \circ (X - M)\|_1. \tag{3.3}$$

Proof. We begin by eliminating the inner maximization in (3.2). Since $\|X - C\|_F^2 = \sum_{ij}(x_{ij} - c_{ij})^2$, this maximization separates into scalar problems of the form

$$\max_{c_{ij}} \quad (x_{ij} - c_{ij})^2, \quad \text{s.t. } l_{ij} \leqslant c_{ij} \leqslant u_{ij} \ \text{ for } 1 \leqslant i, j \leqslant n.$$

A brief calculation shows that the optimal value of c_{ij} is given by

$$c_{ij} = \begin{cases} u_{ij} & \text{if } x_{ij} < m_{ij}, \\ l_{ij} & \text{if } x_{ij} \geqslant m_{ij}. \end{cases}$$

This solution can also be compactly written as

$$c_{ij} = m_{ij} - \operatorname{sgn}(x_{ij} - m_{ij}) d_{ij},$$

using which the value of the objective function becomes

$$\sum_{ij}(x_{ij} - c_{ij})^2 = \sum_{ij}[(x_{ij} - m_{ij})^2 + 2|x_{ij} - m_{ij}|d_{ij} + d_{ij}^2].$$

Dropping constant terms and using matrix notation, we obtain problem (3.3). □

Problem (3.3) is convex, and well solvable by a variety of techniques. But unlike the ordinary NCM problem, it is nondifferentiable, which makes it harder to solve. Therefore, in Section 4, we present an efficient iterative technique to solve this problem.

3.2 *Exploratory correlation matrix nearness*

While in general we might prefer conservative robust models, when exploring different uncertainty models (different boxes $\mathscr{B}$ in our case) it may be more natural to first compute the 'best' matrices in the intersection of $\mathscr{B}$ and $\mathscr{C}_n$. If this intersection is empty, then we would like to find the point in $\mathscr{B}$ closest to $\mathscr{C}_n$. To compute these matrices we may solve

$$\min_{X \in \mathscr{C}_n} \min_{C \in \mathscr{B}} \quad \|X - C\|_F^2. \tag{3.4}$$

Typically, (3.4) admits multiple solutions: there can be several pairs (X^*, C^*) that minimize the objective functional. To ensure uniqueness, we regularize (3.4) by adding a penalty based on departure from an

initial guess G and solve

$$\min_{X\in\mathscr{C}_n}\min_{C\in\mathscr{B}} \quad \|X-C\|_F^2+\lambda\|X-G\|_F^2. \tag{3.5}$$

We refer to (3.5) as the *Exploratory Nearest Correlation Matrix* (E-NCM) problem.

The penalty term in (3.5) has the following impact. As $\lambda\to\infty$, it turns the problem into the projection of G onto $\mathscr{C}_n$, i.e., to the original NCM problem. If $\lambda=0$, the problem may have multiple solutions, while for $\lambda>0$, the problem has a unique 'explorative' solution nearest to the input guess G. This means that λ controls the region of exploration. If no initial guess G is available, we can set $G=0$, so the penalty term reduces to classical Tikhonov regularization.

The solution of (3.4) is independent of the order of minimization, which can be swapped, as we are finding the two nearest points of both sets $\mathscr{C}_n$ and $\mathscr{B}$. Nevertheless, this formulation allows us to recast (3.5) into a form that is easier to analyse, by eliminating the inner minimization.

LEMMA 3.3 (E-NCM–convex formulation) Let M and D be defined as in Lemma 3.2. Then, (3.5) is equivalent to the following convex optimization problem

$$\min_{X\in\mathscr{C}_n} \quad \|[|X-M|-D]_+\|_F^2+\lambda\|X-G\|_F^2, \tag{3.6}$$

where the projection $[\cdot]_+$ and absolute value $|\cdot|$ operators are applied elementwise.

Proof. As the inner minimization in (3.5) is separable, it suffices to consider

$$\min_{c_{ij}} \quad (x_{ij}-c_{ij})^2, \quad \text{s.t. } l_{ij}\leqslant c_{ij}\leqslant u_{ij}, \quad \text{for } 1\leqslant i,j\leqslant n,$$

for which the optimal c_{ij} is easily seen to be

$$c_{ij}=\max\{l_{ij},\min\{u_{ij},x_{ij}\}\}.$$

Substituting this value into the objective and simplifying, the objective over C becomes

$$\sum_{ij}(x_{ij}-c_{ij})^2=\sum_{ij}[|x_{ij}-m_{ij}|-d_{ij}]_+^2,$$

which combined with (3.5) immediately yields the convex formulation (3.6). Moreover, if $\lambda>0$, the problem is strictly convex, which ensures uniqueness. □

3.2.1 *Matrix completion* The E-NCM framework also allows us to consider matrix completion problems as a special case. Suppose that we are given a matrix C with missing entries and we are required to fill-in these entries to complete C to a correlation matrix. In general, such a completion might not exist because the entries of C might enforce constraints that do not intersect with $\mathscr{C}_n$. However, we can still solve E-NCM to tackle this case in a natural way: we set the uncertainty bounds $l_{ij}=u_{ij}=c_{ij}$ for the observed entries, and set $l_{ij}=-1$ and $u_{ij}=1$ for the unknown entries.

If we insist on a unique solution, on one hand we can regularize towards $G=0$ by choosing a trade-off parameter $\lambda>0$. Although this turns into classical Tikhonov regularization, it expresses a preference towards correlation matrices with low ℓ_2 norms. This can be convenient or not, depending on the application framework. On the other hand, if we have some actual meaningful initial guess $G\neq 0$, and we use a small enough λ, then E-NCM approximates the NCM problem with some fixed entries.

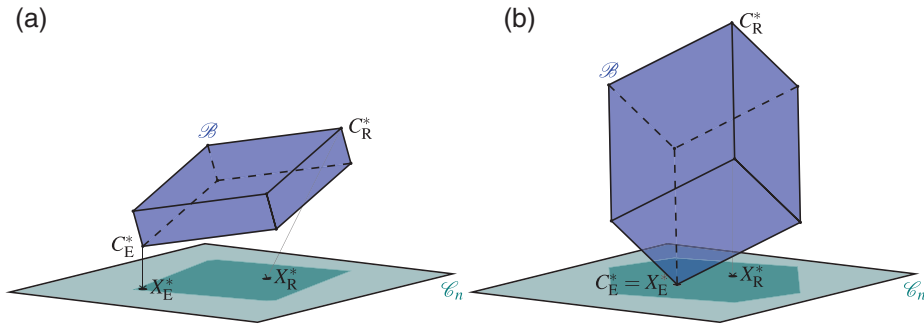

FIG. 1. Schematic representation of the two proposed variants for the NCM problem. In (a) the conservative R-NCM is preferable, while in (b) E-NCM finds the solution that lies inside the box.

3.3 *Discussion*

Figure 1 shows a simplified illustration of R-NCM and E-NCM, where the set $\mathscr{C}_n$ is represented by a plane. The projection of $\mathscr{B}$ over $\mathscr{C}_n$, denoted $\mathrm{P}_{\mathscr{C}_n}(\mathscr{B})$, is depicted as a shadow region (this projection can be interpreted as the set of possible solutions, because for any observation matrix in $\mathscr{B}$ its NCM belongs to this set). E-NCM minimizes the distance to the uncertainty set, and therefore yields the pair of points $X_{\mathrm{E}}^* \in \mathscr{C}_n$ and $C_{\mathrm{E}}^* \in \mathscr{B}$ whose distance is minimum. In contrast, R-NCM minimizes the worst-case scenario, so the solution is given by the point $X_{\mathrm{R}}^* \in \mathscr{C}_n$ such that the distance to the furthest point $C_{\mathrm{R}}^* \in \mathscr{B}$ (the worst-case distance) is minimum.

In Fig. 1(a), the solution provided by E-NCM is close to the border of $\mathrm{P}_{\mathscr{C}_n}(\mathscr{B})$, while R-NCM gives a solution close to the centre. In contrast, Fig. 1(b) shows a situation where the solution of E-NCM and R-NCM are at about the same distance to the border of $\mathrm{P}_{\mathscr{C}_n}(\mathscr{B})$. In this case the sets $\mathscr{B}$ and $\mathscr{C}_n$ intersect, therefore E-NCM finds a solution contained in $\mathscr{B}$.

4. Optimization

In the previous sections, we have introduced two uncertainty-based versions of NCM, and formulated them as convex optimization problems. Although a general-purpose convex optimization solver could be used, it is valuable to derive customized algorithms for the new formulations. In fact, a carefully designed optimization procedure is known to help significantly even for the standard NCM problem (Higham, 2002; Qi & Sun, 2006, 2010; Borsdorf & Higham, 2010). In order to develop such customized algorithms, we need to address two nontrivial issues: (i) we need to project onto the set $\mathscr{C}_n$; and (ii) we need to handle nondifferentiability of the objective functionals. We deal with both these concerns by introducing a unified optimization framework.

We begin by recalling some convex analysis preliminaries; for more details, please refer to Bauschke & Combettes (2011).

4.1 *Background*

Let $\mathscr{X}$ denote a subset of a Euclidean space $\mathscr{E}$. A function $f : \mathscr{X} \to \mathbb{R} \cup \{-\infty, +\infty\}$ is called *lower semicontinuous* (lsc) if for every $x \in \mathscr{X}$, whenever a sequence $(x_k) \to x$, it holds that

$$f(x) \leqslant \liminf_k f(x_k). \tag{4.1}$$

We denote the set of lsc convex functions over $\mathscr{X}$ by $\Gamma_0(\mathscr{X})$. An important example of an lsc convex function is the *indicator function* of a closed convex set $\mathscr{C}$, defined as the function $\delta_{\mathscr{C}} : \mathscr{C} \to \mathbb{R} \cup \{+\infty\}$ such that

$$\delta_{\mathscr{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathscr{C}, \\ +\infty & \text{otherwise.} \end{cases} \tag{4.2}$$

For two lsc convex functions f and g, the *infimal convolution* is defined as

$$(f \,\square\, g)(x) := \inf_{y \in \mathscr{X}} (f(y) + g(x - y)), \tag{4.3}$$

of which one of the most important special cases is the infimal convolution of a convex function with the squared Euclidean norm, which yields the *Moreau envelope* (Moreau, 1962).

PROPOSITION 4.1 Let $f \in \Gamma_0(\mathscr{X})$ and let $\gamma > 0$. The *Moreau envelope* of f indexed by γ is defined as

$$E_f^{\gamma}(x) := \left(f \,\square\, \frac{1}{2\gamma} \|\cdot\|^2 \right)(x). \tag{4.4}$$

The Moreau envelope (4.4) is convex, real-valued and continuous.

Proof. See (Bauschke & Combettes, 2011, Prop. 12.15). □

The Moreau envelope leads us to an operator crucial to our algorithms.

DEFINITION 4.2 (Proximity operator) Let $f \in \Gamma_0(\mathscr{X})$ and $y \in \mathscr{E}$. Then, $\operatorname{prox}_f(y)$ is the unique point in $\mathscr{X}$ that satisfies $E_f^1(y) = \min_{x \in \mathscr{X}} (f(x) + \frac{1}{2}\|x - y\|^2)$, that is,

$$\operatorname{prox}_f(y) := \operatorname*{argmin}_{x \in \mathscr{X}} f(x) + \frac{1}{2}\|x - y\|^2, \tag{4.5}$$

and the operator $\operatorname{prox}_f : \mathscr{X} \to \mathscr{X}$ is called the *proximity operator* of f.

Proximity operators play a very important role in nonsmooth convex optimization, in part because they greatly generalize orthogonal projection. Indeed, if f is the indicator function $\delta_{\mathscr{C}}$ of a closed convex set $\mathscr{C}$, then prox_f reduces to the orthogonal projection operator onto $\mathscr{C}$. The importance of the proximity operator in convex optimization partially stems from the following fact.

PROPOSITION 4.3 For any $f \in \Gamma_0(\mathscr{X})$ and $\gamma > 0$, a point $x \in \mathscr{X}$ is a minimizer of f if and only if

$$x = \operatorname{prox}_{\gamma f}(x). \tag{4.6}$$

The optimality condition (4.6) can be readily exploited to generate a sequence that approximates a minimizer of f:

$$x_{k+1} = \operatorname{prox}_{\gamma f}(x_k).$$

This iterative procedure is nothing, but Rockafellar's famous *proximal point algorithm* (Rockafellar, 1976), which generates a sequence convergent to a minimizer of f under fairly mild conditions. Unfortunately, in the form given above, the proximal point algorithm is often impractical, since computing the

proximity operator of the objective functional is generally difficult. Let us therefore recall a powerful optimization technique that is particularly effective in our case: *Douglas–Rachford splitting*.

4.2 *Douglas–Rachford splitting*

In general, computing proximity operators of generic objective functions can be difficult. But often the objective function can be decomposed as a sum of two parts, and each of these parts admits efficient proximity computation. The Douglas–Rachford (DR) splitting method takes advantage of such a possibility by 'splitting' the optimization task into a series of proximity computations. DR splitting dates back to a method for solving matrix equations of the form $Ax + Bx = y$ for positive definite A and B (Douglas & Rachford, 1956). Since then, this method has been substantially generalized, and is currently enjoying substantial renewed interest—see Combettes & Pesquet (2011) for more details and numerous references.

The starting point for DR is the optimization problem

$$\min_{x\in\mathbb{R}^n} \quad f_1(x) + f_2(x), \tag{4.7}$$

where f_1 and $f_2 \in \Gamma_0(\mathbb{R}^n)$ are functions that satisfy the conditions (which ensure existence of a solution):

$$\operatorname{relint}(\operatorname{dom} f_1) \cap \operatorname{relint}(\operatorname{dom} f_2) \neq \emptyset,$$
$$f_1(x) + f_2(x) \to \infty \quad \text{as } \|x\|_2 \to \infty.$$

For $\gamma > 0$, a solution x to (4.7) satisfies the following optimality conditions (Combettes & Pesquet, 2011)

$$x = \operatorname{prox}_{\gamma f_2} y, \quad \operatorname{prox}_{\gamma f_2} y = \operatorname{prox}_{\gamma f_1}(2\operatorname{prox}_{\gamma f_2} y - y). \tag{4.8}$$

These conditions lie at the heart of the DR algorithm, which is presented as Algorithm 1. Essentially, the DR algorithm iteratively enforces the optimality conditions (joint fixed-point conditions) (4.8). The sequence x^t generated by Algorithm 1 is guaranteed to converge asymptotically to a solution of problem (4.7) provided that such solution exists, see Bauschke & Combettes (2011).

Algorithm 1: Douglas-Rachford Splitting.

Input: $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$;
Output: $x^* := \lim_{t\to\infty}(x^t) = \operatorname{argmin}[f_1(x) + f_2(x)]$;
Fix $\eta \in (0,1)$, $\gamma > 0$, $y^0 \in \mathbb{R}^n$;
for $t = 0, 1, \ldots$ **do**
  % Prox f_2.
  $x^t \leftarrow \operatorname{prox}_{\gamma f_2}(y^t)$;
  % Step-size.
  Set $\lambda^t \in [\eta, 2-\eta]$;
  % Prox f_1.
  $y^{t+1} \leftarrow y^t + \lambda^t(\operatorname{prox}_{\gamma f_1}(2x^t - y^t) - x^t)$;
end

4.3 *DR for NCM with uncertainty*

Now we turn to showing how to apply the DR idea to our robust NCM formulations. Recall that the objective functionals for the R-NCM and E-NCM problems are given by

$$f_R(X) := \|X - M\|_F^2 + 2\|D \circ (X - M)\|_1,$$
$$f_E(X) := \lambda\|X - G\|_F^2 + \|[|X - M| - D]_+\|_F^2,$$

where the subscripts R and E indicate 'robust' and 'explorative', respectively. By introducing the indicator function of the correlation matrix set, problems (3.3) and (3.6) can be rewritten as

$$\min_X \quad f_R(X) + \delta_{\mathscr{C}_n}(X), \tag{4.9}$$
$$\min_X \quad f_E(X) + \delta_{\mathscr{C}_n}(X). \tag{4.10}$$

The key question now is: how should we split the objective functions of (4.9) and (4.10) before we can invoke DR? Simply decomposing the task into proximity over f_R (f_E) and $\delta_{\mathscr{C}_n}$ is *not* practical: the proximity operator for the indicator function requires solving the basic NCM projection problem, and in the course of solving the overall problems (4.9) and (4.10) we might have to repeat this projection many times. We wish to avoid this undue cost[1] while specializing Algorithm 1 to solve (4.9) and (4.10).

In light of this discussion, we propose a more practical splitting. Let the first function be the indicator of $\mathscr{S}_+^n$, and let the second be the sum of the appropriate convex function (f_R or f_E) and the indicator function of S_D^n. Formally, we set

$$f_1 = \delta_{\mathscr{S}_+^n},$$
$$f_2 = \begin{cases} f_R + \delta_{S_D^n} & \text{for R-NCM,} \\ f_E + \delta_{S_D^n} & \text{for E-NCM.} \end{cases}$$

This splitting is valid since $\mathscr{C}_n = \mathscr{S}_+^n \cap S_D^n$, whereby

$$\delta_{\mathscr{C}_n} = \delta_{\mathscr{S}_+^n} + \delta_{S_D^n}.$$

The proximity operator of f_1 is nothing, but the orthogonal projection onto $\mathscr{S}_+^n$. Such projection can be performed by using the eigenvector decomposition $X = Q_X \Lambda_X Q_X^{-1}$, truncating the negative eigenvalues to 0, and then recomposing the matrix with the non-negative eigenvalues, namely

$$\text{Proj}_{\mathscr{S}_+^n}(X) = Q_X [\Lambda_X]_+ Q_X^{-1}.$$

The proximity operator of f_2 can be computed by first evaluating the proximity operator of f_R or f_E, and then projecting the result onto S_D^n. Expressions for the proximity operator of f_R or f_E are provided below; projection onto S_D^n can be computed by simply setting the diagonal entries to 1.

[1] There is also a theoretical reason for this choice: the basic DR algorithm requires proximity operators to be *exact*, whereas projection onto $\mathscr{C}_n$ can be computed only *inexactly*, as it lacks a closed-form solution.

A brief calculation shows that the proximity operator of f_R with parameter γ is given by

$$\operatorname{prox}_{\gamma f_R}(x) = \begin{cases} \dfrac{x + 2\gamma u}{1 + 2\gamma} & \text{if } x \leqslant m - 2\gamma d, \\ m & \text{if } m - 2\gamma d \leqslant x \leqslant m + 2\gamma d, \\ \dfrac{x + 2\gamma l}{1 + 2\gamma} & \text{if } m + 2\gamma d \leqslant x. \end{cases} \tag{4.11}$$

Similarly, a short derivation shows that the proximity operator of f_E with parameter γ is given by

$$\operatorname{prox}_{\gamma f_E}(x) = \begin{cases} \dfrac{x + 2\gamma l + 2\gamma\lambda g}{1 + 2\gamma + 2\gamma\lambda} & \text{if } x \leqslant l + 2\gamma\lambda(l - g), \\ \dfrac{x + 2\gamma\lambda g}{1 + 2\gamma\lambda} & \text{if } l + 2\gamma\lambda(l - g) \leqslant x \leqslant u + 2\gamma\lambda(u - g), \\ \dfrac{x + 2\gamma u + 2\gamma\lambda g}{1 + 2\gamma + 2\gamma\lambda} & \text{if } u + 2\gamma\lambda(u - g) \leqslant x. \end{cases} \tag{4.12}$$

The last remaining detail needed to complete our algorithm is the stopping criterion. For that, we use absolute normalized change

$$\frac{1}{n}\|X^{(t)} - X^{(t-1)}\|_F < \varepsilon_{\text{Stop}},$$

where $\varepsilon_{\text{Stop}}$ is a small constant. This criterion coincides (up to a constant) with the residual of the nonlinear fixed-point equation iterated by DR. The overall algorithms are described in Algorithm 2 for R-NCM and in Algorithm 3 for E-NCM.

Both algorithms can be extended to a W-weighted version of problems (3.2) and (3.5) because the proximity operators are applied elementwise. Thus, the weighted proximity operators of a component x_{ij} with a weight w_{ij} and a step $\hat{\gamma}$ can be computed using the unweighted proximity operators of (4.11) and (4.12) with step $\gamma = \hat{\gamma} w_{ij}$. The rest of the algorithms will remain the same. However, an effect that is similar to the presence of weights can be obtained by varying the widths of the box constraints.

5. Experiments

In this section, we present several experiments to illustrate the behaviour of our new NCM variants; we also compare their computational costs with that of the original NCM problem.

5.1 *Computational time*

Figure 2 compares the computational time and the number of iterations for R-NCM, E-NCM and the original NCM problem. Since E-NCM requires choosing a regularization parameter λ, we show results with three values at different scales, namely 10^{-3}, 10^{-2} and 10^{-1}.

For the experiments, we generate random correlation matrices using the `gallery` function of Matlab with some eigenvalues close to zero; these matrices are then perturbed with random noise and truncated to ensure that their entries lie in $[-1, 1]$. More precisely, we generate[2] the matrices in a manner similar to Qi & Sun (2011).

[2] The Matlab code is: `G=10.^[-4:4/(n-1):0]; G=gallery('randcorr',n*G/sum(G)); G=0.9*G+0.1*(2*rand(size(G))-1); G=(G+G')/2.`

Algorithm 2: Algorithm for solving R-NCM using DR.

Input: $L, U \in \mathscr{S}^n$; % Define $\mathscr{B} = \{X : L \leqslant X \leqslant U\}$.
Output: $X^{(t)} \simeq \min_{X \in \mathscr{C}_n} \max_{C \in \mathscr{B}} \|X - C\|_F^2$;
$M \leftarrow \frac{L+U}{2}$;
Set $\eta \in (0, 1)$, $\gamma > 0$, $Y^{(0)} \in \mathscr{S}^n$;
$t \leftarrow 0$; $\epsilon \leftarrow \infty$; $X^{(0)} = k\mathbf{1}_n$, with $k \gg 1$; % Ensure first iterations.
while $\epsilon \geqslant \varepsilon_{\text{Stop}}$ **do**
  % Prox f_2.
  $X^{(t)} \leftarrow \text{ProxRobust}(Y^{(t)}; L, U; \gamma)$; % Proximity operator (4.11).
  $X_{ii}^{(t)} \leftarrow 1, i = 1, \ldots, n$; % Projection over S_D.
  % Step-size.
  Set $\lambda^{(t)} \in [\eta, 2 - \eta]$;
  % Prox f_1.
  $T \leftarrow 2X^{(t)} - Y^{(t)}$;
  $Q_T \Lambda_T Q_T^{-1} \leftarrow T$; % Eigen decomposition.
  $T \leftarrow Q_T [\Lambda_T]_+ Q_T^{-1}$; % Projection over $\mathscr{S}_+^n$.
  $Y^{(t+1)} \leftarrow Y^{(t)} + \lambda^{(t)}(T - X^{(t)})$; % Update.
  % Stopping condition.
  $t \leftarrow t + 1$; $\epsilon \leftarrow \frac{1}{n}\|X^{(t)} - X^{(t-1)}\|_F^2$;
end

Algorithm 3: Algorithm for solving E-NCM using DR.

Input: $L, U, G \in \mathscr{S}^n$, $\lambda \in \mathbb{R}$; % Define $\mathscr{B} = \{X : L \leqslant X \leqslant U\}$.
Output: $X^{(t)} \simeq \min_{X \in \mathscr{C}_n} \min_{C \in \mathscr{B}} \|X - C\|_F^2 + \lambda \|X - G\|_F^2$;
Set $\eta \in (0, 1)$, $\gamma > 0$, $Y^{(0)} \in \mathscr{S}^n$;
$t \leftarrow 0$; $\epsilon \leftarrow \infty$; $X^{(0)} = k\mathbf{1}_n$, with $k \gg 1$; % Ensure first iterations.
while $\epsilon \geqslant \varepsilon_{\text{Stop}}$ **do**
  % Prox f_2.
  $X^{(t)} \leftarrow \text{ProxExploratory}(Y^{(t)}; L, U, G; \gamma, \lambda)$; % Proximity operator (4.12).
  $X_{ii}^{(t)} \leftarrow 1, i = 1, \ldots, n$; % Projection over S_D.
  % Step-size.
  Set $\lambda^{(t)} \in [\eta, 2 - \eta]$;
  % Prox f_1.
  $T \leftarrow 2X^{(t)} - Y^{(t)}$;
  $Q_T \Lambda_T Q_T^{-1} \leftarrow T$; % Eigen decomposition.
  $T \leftarrow Q_T [\Lambda_T]_+ Q_T^{-1}$; % Projection over $\mathscr{S}_+^n$.
  $Y^{(t+1)} \leftarrow Y^{(t)} + \lambda^{(t)}(T - X^{(t)})$; % Update.
  % Stopping condition.
  $t \leftarrow t + 1$; $\epsilon \leftarrow \frac{1}{n}\|X^{(t)} - X^{(t-1)}\|_F^2$;
end

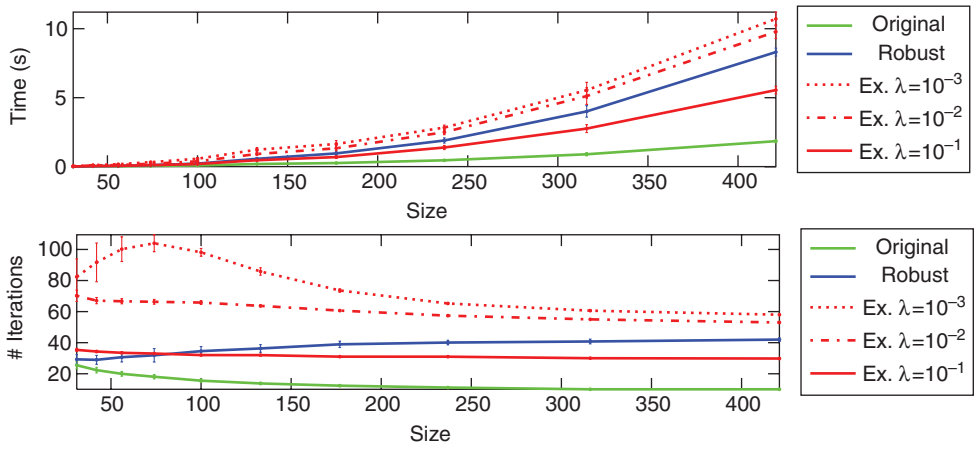

FIG. 2. Computational time and number of iterations with respect to the size for the original NCM problem, R-NCM and E-NCM.

We centre the uncertainty box over the corresponding perturbed correlation matrix, so that its entry-wise width is given by a uniformly distributed random uncertainty matrix, with about 20% of its entries equal to zero (so 20% is considered fixed). The box is then truncated to lie into $[-1, 1]^n$. For each size, we generate 50 random examples. The stopping threshold is $\varepsilon_{\text{Stop}} = 10^{-4}$.

Unsurprisingly, the original NCM is the fastest problem to solve as it is the least restrictive. R-NCM attains the middle position, while for E-NCM, the smaller the parameter λ, the slower it becomes. This happens since, for small λ, the gradient of the regularization term contributes a small amount, which makes the problem depend more on the nonsmooth part. When $\lambda \gg 0$, E-NCM tends to the original NCM problem, which is much easier to solve. The same remarks apply to the number of iterations, which seems to stabilize as the size grows, again with NCM requiring the fewest iterations and E-NCM with small λ, the most.

5.2 *Evolution of the residual*

Recall that we use the DR residual as the stopping criterion of the algorithms. Figure 3 plots this residual versus the number of iterations for two matrices of sizes 50×50 and 250×250.

As before, the easiest problem is NCM, whose residual achieves the machine epsilon limit in both cases, while R-NCM lies between NCM and E-NCM in speed. Note that although there is no reason to expect monotonic descent, the results show a fairly stable descent, except for R-NCM for the smaller matrix. Nevertheless, there does not seem to be any severe ill effects of fluctuation.

5.3 *Application to NCM with a few prescribed entries*

Consider the following 5×5 NCM problem from Qi & Sun (2010). This example requires computing the NCM to the matrix

$$G = \begin{pmatrix} 1.00 & \mathbf{-0.50} & -0.30 & \mathbf{-0.25} & \mathbf{-0.70} \\ & 1.00 & \mathbf{0.90} & 0.30 & 0.70 \\ & & 1.00 & 0.25 & 0.20 \\ & & & 1.00 & 0.75 \\ & & & & 1.00 \end{pmatrix},$$

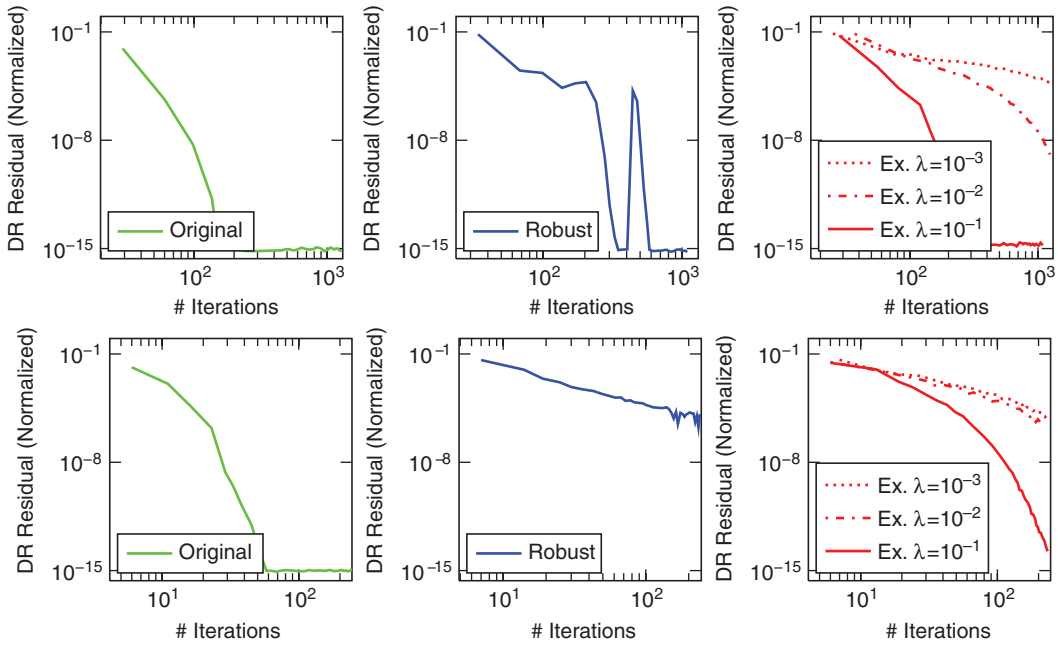

FIG. 3. Evolution of the residual for a 50 × 50 matrix (upper row) and a 250 × 250 matrix (lower row).

such that the bold entries remain fixed. This is equivalent to the following NCM problem with additional equality constraints

$$\min_{X \in \mathscr{C}_n} \quad \|X - G\|_2^2$$
$$\text{s.t.} \quad X_{12} = G_{12}, \quad X_{14} = G_{14}, \quad X_{15} = G_{15}, \quad X_{23} = G_{23}.$$

We may also tackle this constrained NCM problem as an instance of E-NCM. The lower and upper bound matrices L and U are set to -1 and 1 for the unknown (uncertain) entries, and to the corresponding values of G for the fixed (prescribed) ones. Thus,

$$L = \begin{pmatrix} 1.00 & -0.50 & -1.00 & -0.25 & -0.70 \\ & 1.00 & 0.90 & -1.00 & -1.00 \\ & & 1.00 & -1.00 & -1.00 \\ & & & 1.00 & -1.00 \\ & & & & 1.00 \end{pmatrix}, \quad U = \begin{pmatrix} 1.00 & -0.50 & +1.00 & -0.25 & -0.70 \\ & 1.00 & 0.90 & +1.00 & +1.00 \\ & & 1.00 & +1.00 & +1.00 \\ & & & 1.00 & +1.00 \\ & & & & 1.00 \end{pmatrix}.$$

The corresponding solution obtained with parameters $\lambda = 10^{-4}$ and $\varepsilon_{\text{Stop}} = 10^{-10}$ is

$$X^* = \begin{pmatrix} 1.0000 & -0.5000 & -0.2830 & -0.2500 & -0.7000 \\ & 1.0000 & 0.9000 & 0.3391 & 0.6134 \\ & & 1.0000 & 0.2179 & 0.2710 \\ & & & 1.0000 & 0.7198 \\ & & & & 1.0000 \end{pmatrix}.$$

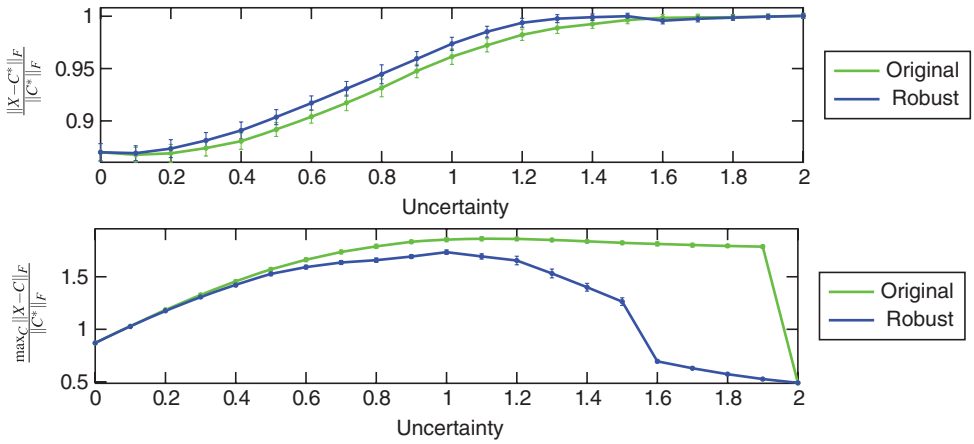

FIG. 4. Completion error and maximum (robust) error as functions of the uncertainty (width of $\mathscr{B}$).

This solution is the same as the one reported in Qi & Sun (2010). But the point worth noting here is that our algorithm does *not impose* any hard constraints on the problem. That is, if there were no correlation matrix with the prescribed fixed entries (as it would happen if intersection between the box and $\mathscr{C}_n$ is empty), our algorithm would still find a correlation matrix near to the box.

5.4 *Experiments with matrix completion*

Figure 4 depicts the error obtained when tackling the problem of completing matrices.

A random matrix is generated as the centre of an uncertainty box, whose width is varied from zero to two (at its maximum width, the box covers all the possible correlation matrices). Inside each box, a random matrix C^* is generated as the 'real' unknown observation matrix. The original NCM problem is solved by projecting the medium point of the box. It is compared with the solution of R-NCM. Two different error measures are presented. First, the distance to the observation matrix C^* and second, the maximum distance to any point of the box. Note that the latter is just the objective functional of R-NCM (the error in the worst-case scenario).

Obviously, R-NCM achieves the best worst-case error, as it is precisely the objective of the R-NCM minimization problem. Moreover, considering the distance to the observation matrix into the box C^*, R-NCM also performs better than mere projection of the middle point when the uncertainty is almost maximum. When the level of uncertainty is moderate, R-NCM is only slightly worse. Therefore, R-NCM appears to provide the most reasonable estimation of an unknown matrix into a known set of uncertainty intervals (box), while also protecting against the worst-case scenario.

6. Discussion and extensions

We presented two major extensions of the original NCM problem of Higham (2002) by allowing uncertainty in the input data. Specifically, we considered an uncertainty model specified by a box (set of intervals) that contain the data matrix.

Within this model of uncertainty, our first new NCM variant is based on the theory of robust optimization, which involves optimization of the worst-case scenario. That is, it seeks to minimize the

maximum distance of the solution to any possible data matrix from within the uncertainty region. Our experiments support the robust approach, showing it to be useful even when minimizing the distance to an unknown data matrix; this setup also allows us to apply the method to correlation matrix completion.

Our second variant of NCM looks for the minimum distance between the box and the set of correlation matrices. This problem can be used as an approximation to the NCM problem with fixed entries, and it can solve the NCM problem using an ϵ-insensitive error induced by the box.

We formulate both variants as convex optimization problems, which we subsequently solve efficiently by designing an iterative method based on DR operator splitting. Several inexact second-order methods (usually generalizations of the Newton projection method) have been designed for the original NCM problem. A possible extension of this work would be to define such second-order methods for the presented variants.

A more theoretical investigation might entail studying different models of uncertainty that would further enrich the setting introduced in this paper. In fact, the only requirement would be to compute the proximity operator of the maximum distance to the uncertainty set for the robust approach, and of the minimum distance in the case of the exploratory one. Nevertheless, even for simple uncertainty sets such as the ellipsoid (which is thoroughly used in robust optimization), these operations do not have a closed-form solution, and thus they would require an inner optimization algorithm.

We close by mentioning that the robust variants that we presented, yield as special cases many of the previously studied variants of NCM problems. Also, since the new variants are easy to interpret, they might prove helpful to an end user. In particular, E-NCM can be used to tackle the problem of matrix completion, or NCM with fixed entries, as shown in Section 5.3. The key consequence of our proposal is that it imposes the constraints in a soft manner.

Funding

Spain's grant (TIN2010-21575-C02-01) and UAM–ADIC Chair for Machine Learning to C.A.

References

Bauschke, H. H. & Combettes, P. L. (2011) *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Berlin: Springer.

Ben-Tal, A., El Ghaoui, L. & Nemirovski, A. S. (2009) *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton, NJ: Princeton University Press.

Bertsimas, D. & Thiele, A. (2006) Robust and data-driven optimization: modern decision making under uncertainty. *Tutorials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*. Catonsville, MD: Institute for Operations Research and the Management Sciences, pp. 95–122.

Bhansali, V. & Wise, M. B. (2001) Forecasting portfolio risk in normal and stressed markets. *J. Risk*, **4**, 91–106.

Borsdorf, R. & Higham, N. (2010) A preconditioned Newton algorithm for the nearest correlation matrix. *IMA J. Numer. Anal.*, **30**, 94–107.

Combettes, P. & Pesquet, J. C. (2011) Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Berlin: Springer, pp. 185–212.

Dantzig, G. (1955) Linear programming under uncertainty. *Manage. Sci.*, **1**, 197–206.

Douglas, J. & Rachford, H. H. (1956) On the numerical solution of heat conuction problems in two or three space variables. *Trans. Amer. Math. Soc.*, **82**, 421–439.

Gentle, J. E. (2007) *Matrix algebra, theory, computations, and applications in statistics*. Berlin: Springer.

Higham, N. J. (1989) Matrix nearness problems and applications. *Appl. Matrix Theory*. Oxford, UK: Oxford University Press, pp. 1–27.

Higham, N. J. (2002) Computing the nearest correlation matrix: a problem from finance. *IMA J. Numer. Anal.*, **22**, 329–343.

Li, G., Ma, K. C. & Pong, T. K. (2013) Robust least square semidefinite programming with applications to correlation stress testing. *Optimization Online Preprints.*

Moreau, J. J. (1962) Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Sér. A Math.*, **255**, 2897–2899.

Qi, H. & Sun, D. (2006) A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM J. Matrix Anal. Appl.*, **28**, 360–385.

Qi, H. & Sun, D. (2010) Correlation stress testing for value-at-risk: an unconstrained convex optimization approach. *Comput. Optim. Appl.*, **45**, 427–462.

Qi, H. & Sun, D. (2011) An augmented lagrangian dual approach for the H-weighted nearest correlation matrix problem. *IMA J. Numer. Anal.*, **31**, 491–511.

Rockafellar, R. T. (1976) Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, **14**, 877–989.

Shapiro, A., Dentcheva, D. & Ruszczynski, A. (2009) *Lectures on Stochastic Programming: Modeling and Theory.* Philadelphia, PA: Society for Industrial and Applied Mathematics and Mathematical Programming Society.