Sequential convex relaxation for convex optimization with bilinear matrix equalities

Doelman, Reinier; Verhaegen, Michel

DOI
10.1109/ECC.2016.7810576

Publication date
2016

Document Version
Accepted author manuscript

Published in
Proceedings 2016 European Control Conference (ECC 2016)

Accepted Author Manuscript . Published in: 2016 European Control Conference (ECC)

Sequential convex relaxation for convex optimization with bilinear matrix equalities*

Reinier Doelman[1,2] and Michel Verhaegen[1]

***Abstract*— We consider the use of the nuclear norm operator, and its tendency to produce low rank results, to provide a convex relaxation of Bilinear Matrix Inequalities (BMIs). The BMI is first written as a Linear Matrix Inequality (LMI) subject to a bi-affine equality constraint and subsequently rewritten into an LMI subject to a rank constraint on a matrix affine in the decision variables. The convex nuclear norm operator is used to relax this rank constraint. We provide an algorithm that iteratively improves on the sum of the objective function and the norm of the equality constraint violation. The algorithm is demonstrated on a controller synthesis example.**

I. INTRODUCTION

Many controller synthesis techniques can be efficiently solved using convex optimization and LMIs. Examples include static state-feedback controllers and dynamic output feedback controllers. The crucial last step in deriving these LMIs is usually a convexification step that hinges on substituting decision variables that only appear in products, or more involved transformations, see for example [1], [2], [3]. If no convexification step is known, the control designer is usually left solving a BMI, which in many cases is an NP-hard problem [4]. However, this does not exclude the possibility of efficient algorithms for finding local minima to the optimization problem.

A rough distinction can be made in global and local optimization algorithms, see [5] and the references therein. Many global optimization algorithms are of a Branch and Bound-type, for example [6], [7], but since the problem is NP-hard the computational complexity and run time of these algorithms remains an issue.

Local optimization algorithms can be very intuitive. A dual iteration method, [8], could start from any feasible solution to the BMI, fix one of the bilinear variables and optimize over the others. Subsequently fix the other bilinear variables and and repeat this procedure. There is however no guarantee of convergence of this procedure to a local minimum [9], [6], [10].

Other local optimization algorithms include path-following methods [11], [12] that include a linearization step, and, relevant to this work, rank minimization (rank constrained) methods [13], [14], [15].

The convex relaxation proposed in this paper requires less variables to be introduced to the problem than similar approaches in literature. Furthermore, we propose two iterative algorithms that do not require a feasible solution to the BMI as a starting point.

A. Notation

Let $\mathbb{R}^{n\times m}$ denote the set of real-valued matrices of size $n \times m$ and let $X \in \mathbb{S}^n, X \succeq 0$ denote a symmetric positive semidefinite matrix X of size $n \times n$. $\mathbb{R}^+$ denotes the set of real and positive scalars. $\operatorname{rank}(X)$ denotes the rank of matrix X and $\|X\|_*$ denotes its nuclear norm. X^+ is the notation used for the Moore-Penrose pseudoinverse of a matrix X, and $X^\perp$ denotes the orthogonal complement of X, i.e. $X^\perp X = 0,\ X^\perp X^{\perp T} \succ 0$. In relation to optimization problems, an overbar $\bar{X}$ denotes the optimal value of decision variable X. In symmetric (block) matrices, the symbol $\star$ denotes blocks that can be inferred from symmetry and $\operatorname{sym}(X) = X + X^T$. The use of $(\cdot)$ for function arguments indicates that the arguments can be inferred from the context. Finally, let $\{x^k\}_{k=0,1,\dots}$ denote a sequence of values x^k indexed by k.

B. Problem formulation

The general type of bilinear (bi-affine) optimization problems we are considering are the ones that can abstractly be written as follows. Consider a scalar $\gamma \in \mathbb{R}$, the matrices $A \in \mathbb{R}^{n_a\times n_b},\ B \in \mathbb{R}^{n_c\times n_d}$, and $P \in \mathbb{R}^{n_b\times n_c}$, and a matrix valued function g:

$$\begin{aligned} \underset{\gamma,x,A,B}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & g(\gamma, x, A, B, APB) \succ 0. \end{aligned} \tag{1}$$

Here x denotes a tuple of decision variables (either scalars, vectors or matrices). The matrix P is not a decision variable, but a given matrix originating from the problem at hand. We can substitute the product APB with an additional variable $C \in \mathbb{R}^{n_a\times n_d}$, add an equality constraint and obtain the equivalent problem:

$$\begin{aligned} \underset{\gamma,x,A,B,C}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & g(\gamma, x, A, B, C) \succ 0, \\ & APB = C. \end{aligned} \tag{2}$$

The function $g(\cdot)$ itself is now affine in all decision variables and the bilinearity is contained in the equality constraint.

This formulation expresses many controller synthesis BMIs where $\gamma \in \mathbb{R}^+$ is the (squared) $\mathcal{H}_\infty$ or the (squared) $\mathcal{H}_2$ norm of the resulting system, the function $g(\cdot) \succ 0$ expresses the matrix inequality required for stability and performance guarantee, and A, B and C are decision variables constrained in a bilinear way.

*The work of R. Doelman and M. Verhaegen is sponsored by the European Research Council, Advanced Grant Agreement No. 339681.

[1]Delft University of Technology, The Netherlands.

[2]Corresponding author: r.doelman@tudelft.nl.

For some BMI problems convexification steps exist. One example is where $P = I$, B only appears in products with A, and A is guaranteed invertible. The expression AB is substituted by C and after the optimal solution of the resulting LMI is obtained, the value of B is calculated by $B = A^{-1}C$. Although it is possible to find the globally optimal value of γ after this substitution, one loses the ability to put (structural) constraints on B.

However, sometimes it is possible to formulate the problem in such a way as to still have some structural constraints on B by accepting structural constraints on A and C that would enable convexification (for example the assumption that they are (block) diagonal), but this usually comes at the price of conservatism on the optimal value of γ. In this article we present in Section II an algorithm that iteratively tries to find a solution to problem (1). The algorithm can handle (structural) constraints on decision variables if these originate from the problem, without unnecessarily imposing constraints on other decision variables. In Subsection II-A the bi-affine equality constraint is first rewritten into a rank constraint on a specially constructed matrix M, and this rank constraint is relaxed using the nuclear norm operator. The construction of the matrix M allows for an iterative algorithm that manipulates this matrix at every iteration. In Subsection II-B it is proven that the iterative algorithm defined using this convex relaxation assures that the sum of γ and the norm of the equality constraint violation converges. We relate our work to existing literature in Section III. We provide numerical results in Section IV.

II. ITERATIVE SOLUTION TO BMI (1)

A. Description of the algorithm

To show some of the equalities in the proof of the equivalence of constraints, we need the following lemma on the generalized Schur complement.

Lemma 1 (Carlson [16], generalized Schur complement): Let the matrix X be defined as

$$X = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}. \tag{3}$$

Then $\operatorname{rank}(X) = \operatorname{rank}(X_4) + \operatorname{rank}\left(X_1 - X_2 X_4^+ X_3\right)$ if and only if

$$X_2\left(I - X_4^+ X_4\right) = 0, \tag{4}$$
$$\left(I - X_4 X_4^+\right) X_3 = 0. \tag{5}$$

This lemma is instrumental in proving the following theorem.

Theorem 1: Given any matrices $X \in \mathbb{R}^{n_a \times n_b}$, $Y \in \mathbb{R}^{n_c \times n_d}$ and any full rank square matrices $W_1 \in \mathbb{R}^{n_a \times n_a}$, $W_2 \in \mathbb{R}^{n_d \times n_d}$, define the matrix M:

$$M := \begin{pmatrix} W_1 & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} C + XPY + APY + XPB & (A+X)P \\ P(B+Y) & P \end{pmatrix} \times \begin{pmatrix} W_2 & 0 \\ 0 & I \end{pmatrix}.$$

The following two optimization problems are equivalent.

1) The optimization with bilinear equality constraint:

$$\begin{aligned} \underset{\gamma,x,A,B,C}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & g(\gamma, x, A, B, C) \succ 0, \\ & APB = C. \end{aligned} \tag{6}$$

2) The optimization with rank constraint:

$$\begin{aligned} \underset{\gamma,x,A,B,C}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & g(\gamma, x, A, B, C) \succ 0, \\ & \operatorname{rank}(M) = \operatorname{rank}(P). \end{aligned} \tag{7}$$

Proof: To start, notice that constraint $C = APB$ equals a rank constraint on the difference between C and the product APB: $\operatorname{rank}(C - APB) = 0$. Enforcing this constraint is difficult for two reasons: it is a rank constraint, and the decision variables do not appear affinely in the constraint. However, using Lemma 1, we can rewrite this constraint.

What we know of matrix M is that the conditions of Lemma 1, (4) and (5), are fulfilled, since

$$W_1(A+X)P(I - P^+P) = 0, \quad \text{and}$$
$$(I - PP^+)P(B+Y)W_2 = 0.$$

The generalized Schur complement of P in M is:

$$\begin{aligned} & W_1\left(C + XPY + APY + XPB\right)W_2 \\ & - W_1\left((A+X)P\right)\left(P^+\right)\left(P\left(B+Y\right)\right)W_2 \\ & = W_1\left(C - APB\right)W_2, \end{aligned}$$

so applying Lemma 1 gives us

$$\operatorname{rank}(M) = \operatorname{rank}(P) + \operatorname{rank}\left(W_1\left(C - APB\right)W_2\right).$$

Since W_1, W_2 are square and full rank we have the equivalence $\operatorname{rank}(M) = \operatorname{rank}(P) \iff \operatorname{rank}(C - APB) = 0 \iff APB = C$. ∎

Since the matrix M is affine in all three decision variables, we propose to use the (convex) nuclear norm operator to relax the rank constraint.

Theorem 2: Let $\lambda \in \mathbb{R}^+$ be a regularization parameter. A convex relaxation of problem (1) is the optimization problem

$$\begin{aligned} \underset{\gamma,x,A,B,C}{\text{minimize}} \quad & \gamma + \lambda \|M\|_* \\ \text{subject to} \quad & g(\gamma, x, A, B, C) \succ 0. \end{aligned} \tag{8}$$

If for the (feasible) optimal tuple $(\bar{\gamma}, \bar{x}, \bar{A}, \bar{B}, \bar{C})$ it holds that $\operatorname{rank}\left(\bar{M}\right) = \operatorname{rank}(P)$, then the optimal tuple is a feasible solution of problem (1).

Proof: Let s be the number of singular values of M, and r defined as the number of non-zero singular values (which is the rank) of P:

$$s := \min(n_a + n_b, n_c + n_d),$$
$$r := \operatorname{rank}(P),$$

and let the following notation denote the nuclear norm and the Ky Fan r-norm respectively, $\|M\|_* := \sum_{i=1}^{s} \sigma_i(M)$, $\|M\|_{Fr} := \sum_{i=1}^{r} \sigma_i(M)$, where $\sigma_i(M)$ is the ith largest singular value of M. Since these norms are well known to be convex, we can write the *truncated*

nuclear norm [17] as the difference of two convex functions: $\|M\|_r = \|M\|_* - \|M\|_{Fr}$. Since the rank of a matrix is directly related to the number of non-zero singular values, we can equivalently to problem (7) write the rank constraint as $\|M\|_r = 0$. Instead of enforcing the constraint, we add the truncated nuclear norm to the objective function, weighed by a regularization parameter λ, and after the optimization, check if the constraint $\|M\|_r = 0$ is met:

$$\begin{aligned} \underset{\gamma,x,A,B,C}{\text{minimize}} \quad & \gamma + \lambda \|M\|_r , \\ \text{subject to} \quad & g(\gamma, x, A, B, C) \succ 0. \end{aligned}$$

We now have a Difference of Convex Programming problem (DCP, [7]), since the objective function can be written as

$$\gamma + \lambda \|M\|_r = (\gamma + \lambda \|M\|_*) - \lambda \|M\|_{Fr} . \qquad (9)$$

The convex relaxation is therefore to drop the term $\lambda \|M\|_{Fr}$, and accept the bias this introduces in the solution for the optimal γ. ■

The bias introduced by ignoring the contribution of $\|M\|_{Fr}$ in Equation (9) to the value of the objective function has also been addressed in [17], where the truncated nuclear norm is minimized as a better approximation of the rank function than the nuclear norm.

With the formulation as a DCP problem in (9), it is possible to apply DC algorithms to find minimizing solutions, like the Convex-Concave Procedure in [18]. This optimization method uses the expressions for the (sub)differential of the two convex functions. However, the parameters X, Y in M allow for a different iterative algorithm, for which expressions of (sub)differentials are not necessary. We will present this algorithm next.

The singular values of M constructed from the optimal $\bar{A}, \bar{B}, \bar{C}$ of problem (8) in general depend on the choices of λ, X, Y, W_1, W_2. These parameters do not influence the set of feasible solutions of the constraint $g(\cdot) \succ 0$. However, the choice influences the objective function, and therefore influences the result of optimization of the convex relaxation. The optimization will favour solutions for which $\|M\|_*$ has a low value. Looking at the structure of M, we see that over all possible values of A, B, C, X, Y, W_1, W_2, irrespective of whether these variables satisfy the constraint $g(\cdot) \succ 0$ or not, its minimum nuclear norm equals

$$\min_{A,B,C,X,Y,W_1,W_2} \|M\|_* = \left\| \begin{pmatrix} 0 & 0 \\ 0 & P \end{pmatrix} \right\|_* = \|P\|_* , \qquad (10)$$

and this is attained (when P is square and full rank) for all cases where $A = -X, B = -Y$ and C such that $C = APB$. In this specific case, we also have that $\|M\|_{Fr} = \|P\|_*$ and $\|M\|_r = 0$. The effect of this is that the bias due to ignoring $\|M\|_{Fr}$, is not towards $A = 0, B = 0, C = 0$, but solutions where $A = -X$, $B = -Y$ and $C = APB$. It is important to note that the globally optimal values of the decision variables to BMI (1), say $\bar{A}, \bar{B}, \bar{C}$, might not produce the optimal value of the objective function in (8), even for $X = -\bar{A}$ and $Y = -\bar{B}$.

The freedom to choose the values of X and Y allows for a simple iterative procedure described in Algorithm 1.

Data: Initialization matrices X_a^1, Y_a^1. Matrices $W_1 = I, W_2 = I$. Termination condition variable $\epsilon \in \mathbb{R}^+$, $\epsilon \ll 1$, and regularization parameter $\lambda \in \mathbb{R}^+$.

while $\bar{f}_b^{k-1} - \bar{f}_b^k > \epsilon$ **do**

Solve the first optimization:

$$\begin{aligned} \underset{\gamma_a^k, x_a^k, A_a^k, B_a^k, C_a^k}{\text{minimize}} \quad & \gamma_a^k + \lambda \left\| M_a^k \left(X_a^k, Y_a^k \right) \right\|_* \\ \text{subject to} \quad & g(\gamma_a^k, x_a^k, A_a^k, B_a^k, C_a^k) \succ 0, \\ & P(B_a^k + Y_a^k) = 0; \end{aligned} \qquad (11)$$

Set $X_b^k = -\bar{A}_a^k$ and $Y_b^k = -\bar{B}_a^k$;

Solve the second optimization:

$$\begin{aligned} \underset{\gamma_b^k, x_b^k, A_b^k, B_b^k, C_b^k}{\text{minimize}} \quad & \gamma_b^k + \lambda \left\| M_b^k \left(X_b^k, Y_b^k \right) \right\|_* \\ \text{subject to} \quad & g(\gamma_b^k, x_b^k, A_b^k, B_b^k, C_b^k) \succ 0, \\ & (A_b^k + X_b^k) P = 0; \end{aligned} \qquad (12)$$

Set $X_a^{k+1} = -\bar{A}_b^k$ and $Y_a^{k+1} = -\bar{B}_b^k$;

Calculate $\bar{f}_b^k(\bar{\gamma}_b^k, \bar{A}_b^k, \bar{B}_b^k, \bar{C}_b^k)$.

end

Algorithm 1: Iterative improvement with guaranteed convergence.

The matrix M is initialized with X_a^1, Y_a^1. Care should be taken that the choice of Y_a^1 allows for overlap between the convex set $g(\cdot) \succ 0$ and the convex set $P(B_a^1 + Y_a^1) = 0$, or problem (11) will be infeasible. This can be assured by setting $Y_a^1 = -B$, with B from any feasible tuple (γ, x, A, B, C) to the constraint $g(\cdot) \succ 0$. If the convex set defined by $g(\cdot) \succ 0$ is empty, no solution will exist to problem (1), since the solutions to BMI (1) are a subset of the relaxed problem. The function $\bar{f}_b^k$ in Algorithm 1 is defined as follows. Let f be the following function:

$$f(\gamma, A, B, C) := \gamma + \lambda \|C - APB\|_* + \lambda \|P\|_* ,$$

where the term $\lambda \|P\|_*$ is just a constant, and denote two variations as follows

$$f_\bullet^k := f(\gamma_\bullet^k, A_\bullet^k, B_\bullet^k, C_\bullet^k), \quad \bar{f}_\bullet^k := f(\bar{\gamma}_\bullet^k, \bar{A}_\bullet^k, \bar{B}_\bullet^k, \bar{C}_\bullet^k).$$

Here $\bullet$ denotes the possible presence of a subscript to indicate one of two optimizations in iteration k of Algorithm 1.

B. Proof of convergence of Algorithm 1

Before we provide the theorem on convergence of Algorithm 1 and its proof, we need the following preliminaries.

The function f expresses the sum of the objective variable to be minimized, the norm of the bilinear equality constraint violation and a constant. Note that if we take $W_1 = I, W_2 = I$ and consider the matrix M in Theorem 1 as being a function of A, B, C, X and Y, then we can split M in the following way:

$$\begin{aligned} M(A, B, C, X, Y) = & \begin{pmatrix} C - APB & 0 \\ 0 & P \end{pmatrix} + \\ & \begin{pmatrix} (A + X)P(B + Y) & (A + X)P \\ P(B + Y) & 0 \end{pmatrix}. \end{aligned} \qquad (13)$$

With this insight, we can write

$$\begin{aligned} f &= \gamma + \lambda \left\| C - APB \right\|_* + \lambda \left\| P \right\|_* \\ &= \gamma + \lambda \left\| M\left(A, B, C, -A, -B\right) \right\|_* . \end{aligned}$$

Furthermore, we use the shorthand notations

$$\begin{aligned} M_\bullet^k\left(X, Y\right) &:= M\left(A_\bullet^k, B_\bullet^k, C_\bullet^k, X, Y\right), \\ \bar{M}_\bullet^k\left(X, Y\right) &:= M\left(\bar{A}_\bullet^k, \bar{B}_\bullet^k, \bar{C}_\bullet^k, X, Y\right). \end{aligned}$$

The following lemma is necessary for the proof of convergence.

Lemma 2 (King [19], Theorem 2): Using the matrices $S, P, Q, R \in \mathbb{R}^{n\times n}$, we have

$$\left\| \begin{pmatrix} S & P \\ Q & R \end{pmatrix} \right\|_* \geq \left\| S \right\|_* + \left\| R \right\|_* .$$

Since non-square matrices can be padded with zeros without influencing their nuclear norm, the restriction of S, P, Q, R to square matrices does not restrict generality.

Theorem 3: Assume there exists a tuple (γ, x, A, B, C) of decision variables which are bounded in norm, and that satisfies the constraint $g(\gamma, x, A, B, C) \succ 0$. Using Algorithm 1 to generate a sequence $\{\bar{f}_b^k\}_{k=1,2,\ldots}$, where k denotes the iteration number, the value of $\left|\bar{f}_b^{k+1} - \bar{f}_b^k\right|$ tends to 0 as $k \to \infty$.

Proof: We will show that one iteration of Algorithm 1 cannot increase the value of $\bar{f}_b^k$, i.e.

$$\bar{f}_b^k \geq \bar{f}_a^{k+1} \geq \bar{f}_b^{k+1}. \tag{14}$$

Due to the equality constraints in problems (11) and (12), the second term of M in Equation (13) has values either in the top-right or in the bottom-left block respectively. The proof will continue with the case where $P(B+Y) = 0$ and where $(A+X)P$ appears in the top-right block. The other case is similar.

To prove the first inequality in (14), suppose that the objective function of problem (12) (denoted with subscribts b) of iteration k has a minimum value for the tuple $(\bar{\gamma}_b^k, \bar{x}_b^k, \bar{A}_b^k, \bar{B}_b^k, \bar{C}_b^k)$. For problem (11) (denoted with subscripts a) of iteration $k+1$, we have this tuple as a feasible point, since the matrix inequality and equality constraint are satisfied. Furthermore, *for this tuple* the value of the objective function in problem (11) at iteration $k+1$ will be:

$$\bar{\gamma}_b^k + \lambda \left\| \bar{M}_a^k\left(-\bar{A}_b^k, -\bar{B}_b^k\right) \right\|_* = \bar{f}_b^k.$$

Due to the fact that the optimization is convex, the optimal value of the objective function of problem (11) at iteration $k+1$ will be equal or lower:

$$\bar{f}_b^k \geq \bar{\gamma}_a^{k+1} + \lambda \left\| \bar{M}_a^{k+1}\left(-\bar{A}_b^k, -\bar{B}_b^k\right) \right\|_* .$$

Using Lemma 2, we have the inequality

$$\begin{aligned} &\left\| \bar{M}_a^{k+1}\left(-\bar{A}_b^k, -\bar{B}_b^k\right) \right\|_* = \\ &\left\| \begin{pmatrix} \bar{C}_a^{k+1} - \bar{A}_a^{k+1} P \bar{B}_a^{k+1} & (\bar{A}_a^{k+1} - \bar{A}_b^k)P \\ 0 & P \end{pmatrix} \right\|_* \\ &\geq \left\| \bar{C}_a^{k+1} - \bar{A}_a^{k+1} P \bar{B}_a^{k+1} \right\|_* + \left\| P \right\|_* \\ &= \frac{1}{\lambda}\left(\bar{f}_a^{k+1} - \bar{\gamma}_a^{k+1}\right). \end{aligned} \tag{15}$$

After collecting terms, we can conclude that $\bar{f}_b^k \geq \bar{f}_a^{k+1}$.

As mentioned before, the proof that $\bar{f}_b^{k+1}$ will not be larger than $\bar{f}_a^{k+1}$ after the second optimization in Algorithm 1, is quite similar and we arrive at the inequalities

$$\bar{f}_b^k \geq \bar{f}_a^{k+1} \geq \bar{f}_b^{k+1}.$$

Since the function $f \geq 0$ and we assumed a feasible tuple exists in the convex set defined by $g(\gamma, x, A, B, C) \succ 0$ with variables with bounded norms (and with the mild assumption that we start from this solution by setting $Y_a^1 = -B$), the optimal values in the sequence $\{\bar{f}_b^k\}_{k=1,2,\ldots}$ are bounded from below and above, so must converge and the difference between successive iterates must tend to zero as $k \to \infty$. ■

It is important to note that with a guarantee of convergence of $\{\bar{f}^k\}$ as $k \to \infty$, we cannot conclude that $\left\|C^k - A^k P B^k\right\|_*$ converges to 0 *even if such a solution exists* within the constraints of problem (6). Furthermore, the global optimal solution of problem (6), might not even give the lowest value of the objective function of problems (11) and (12) due to the influence of X and Y on M, and the algorithm might converge to a different point.

The two constraints

$$(A_b^k + X_b^k)P = 0, \quad P(B_a^k + Y_a^k) = 0, \tag{16}$$

play an important role in showing some of the inequalities in the proof of convergence of Algorithm 1. In practice we often observed convergence even without these constraints, see Algorithm 2. We cannot provide a proof of convergence of

Data: Matrices X^1, Y^1. Matrices $W_1 = I, W_2 = I$.
Regularization parameter $\lambda \in \mathbb{R}^+$.
for *fixed number of iterations* **do**
Solve the optimization:

$$\begin{aligned} \underset{\gamma,x,A,B,C}{\text{minimize}} \quad & \gamma^k + \lambda \left\| M^k(X^k, Y^k) \right\|_* \\ \text{subject to} \quad & g(\gamma^k, x^k, A^k, B^k, C^k) \succ 0, \end{aligned}$$

Set $X^{k+1} = -\bar{A}^k$ and $Y^{k+1} = -\bar{B}^k$;
end

Algorithm 2: Simpler iterative algorithm without guaranteed convergence.

the sequence $\{\bar{f}^k\}$ using this algorithm along the same lines of the proof of Theorem 3, since it is possible to construct an example where $\bar{\gamma}^{k+1} + \lambda \left\| \bar{M}^{k+1} \right\|_*$ is less than $\bar{f}^k$, but $\bar{f}^{k+1} > \bar{f}^k$.

III. RELATION TO EXISTING METHODS

When P is square and full rank, the two constraints in Equation (16) respectively fix A_b^k and B_a^k in Algorithm 1: for such P the null space of P consists of only the 0-vector, and the only option is $A_b^k = -X_b^k = \bar{A}_b^{k-1}$ or $B_a^k = -Y_a^k = \bar{B}_a^{k-1}$. Therefore our method bears a resemblance to a dual iteration method. However, we have no requirement of starting from a feasible solution of the BMI. Another difference between dual iteration and Algorithm 1 is that this algorithm minimizes an upper bound for the sum of γ

and the equality constraint violation, and does not explicitly satisfy the equality constraint. Furthermore, when P is not full rank, the value of $A_\bullet^k$ and $B_\bullet^k$ can both be optimized in the same iteration. This holds even more for Algorithm 2, where all decision variables are optimized simultaneously.

In [14] it is shown that the following indefinite quadratic feasibility test (a BMI) with N variables $\hat{x}_i \in \mathbb{R}$, $i = 1, \ldots, N$ and matrices $F_\bullet \in \mathbb{R}^{m \times m}$,

$$F_0 + \sum_{i=1}^{N} \hat{x}_i F_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{x}_i \hat{x}_j F_{ij} \succ 0, \tag{17}$$

is equivalent to a problem formulation with a rank-1 constraint on a positive semidefinite decision variable $\hat{X} \in \mathbb{R}^{N \times N}$. Each bi-affine combination of variables $\hat{x}_i \hat{x}_j$ is substituted with the (i,j)-th element of $\hat{X}$ to obtain the rank constrained problem

$$\begin{aligned} &F_0 + \sum_{i=1}^{N} \hat{x}_i F_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{X}_{ij} F_{ij} \succ 0, \quad \begin{pmatrix} \hat{X} & \hat{x} \\ \hat{x}^T & 1 \end{pmatrix} \succeq 0, \\ &\operatorname{rank}\left(\hat{X}\right) = 1, \end{aligned} \tag{18}$$

where $\hat{x}$ is the vector of $\hat{x}_i$'s. $\hat{X}$ should equal $\hat{x}\hat{x}^T$, even though the possibility of $\hat{x} = 0$ is excluded here due to the rank constraint on $\hat{X}$.

If we would write problem (1) in the form of (18), we see that the method in [14] uses many more additional variables. Suppose on the other hand that problem (1) is a feasibility test of the form

$$g(\hat{x}, \hat{x}\hat{x}^T) \succ 0,$$

then the considered BMIs in [14] can be cast in a form very similar to that of problem (7), but with different rank constraints and with the difference that (7) does not exclude the possibility of $\hat{x} = 0$.

When dealing with matrix products APB, we can formulate this in both formulation (18) and (7), but by introducing extra variables for C and using the substitution described in this paper we will in general need less of these extra decision variables. The Kronecker product $B^T \otimes A$ in essence describes all combinations of matrix elements in the decision variables A and B that appear bi-affinely in problem (1), and their total number is $n_a n_b n_c n_d$. The number of decision variables introduced for a reformulation of problem (1) into the form of (18) would be $(n_a n_b + n_c n_d)(n_a n_b + n_c n_d + 1)/2$, whereas in this paper we introduce $n_a n_d$ extra variables for the matrix C. We expect the nuclear norm operator to use two additional matrices for the implementation, see [20] for details. Since M is of dimensions $(n_a + n_b) \times (n_c + n_d)$, we expect the implementation to introduce $(n_a + n_b)(n_a + n_b + 1)/2 + (n_c + n_d)(n_c + n_d + 1)/2$ additional variables.

The introduction of extra variables to solve problem (1) in our approach will for most sizes of A and B be more efficient, since their number scales quadratically in the dimensions of the matrices and not quadratically in the products of the dimensions.

The method in [15] is similar to the method presented here on several points. However, in [15] there is a prescribed form of the problem, and the method presented here does not have this. Furthermore, the number of additional decision variables introduced in their reformulation grows with the size of the entire matrix inequality, not with the number of bi-affine combinations of decision variables. In a last point of comparison, the bias introduced by the two different methods differ. The relaxation in [15] introduces a bias on the matrix used in that paper to reformulate $g(\gamma, x, A, B, APB) \succ 0$ towards a matrix with low eigenvalues. On the contrary, the bias introduced by the use of $\|M\|_*$ in Theorem 2 will not stimulate decision variables A and B to have small norms, but to be close to X and Y respectively.

IV. EXPERIMENTAL RESULTS

In [21] the problem of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ controller design is analysed and this problem is used as a benchmark problem for the methods described in [11], [12], [22], [23]. On this problem the method of [23] performs best and we will use its performance as comparison.

Consider the continuous time system:

$$\begin{pmatrix} \dot{\xi} \\ z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \mathcal{A} & \mathcal{B}_u & \mathcal{B}_w \\ \mathcal{C}_1 & \mathcal{D}_1 & 0 \\ \mathcal{C}_2 & \mathcal{D}_2 & 0 \end{pmatrix} \begin{pmatrix} \xi \\ u \\ w \end{pmatrix},$$

where $\xi, z_\bullet, u, w$ respectively denote the state, output, input and disturbance. In the benchmark problem we have $\xi \in \mathbb{R}^3, z_1 \in \mathbb{R}^1, z_2 \in \mathbb{R}^1, w \in \mathbb{R}^2$ and $u \in \mathbb{R}^1$. The dimensions of the system matrices are defined accordingly. The mixed $\mathcal{H}_2/\mathcal{H}_\infty$ design problem is to design a state feedback controller $u = \mathcal{K}\xi$ such that the $\mathcal{H}_2$ norm η of the transfer function from w to z_2 of the closed loop system is as low as possible, while the $\mathcal{H}_\infty$ norm γ of the transfer function from w to z_1 does not exceed a prescribed level.

We rewrite the formulation in [12] into that of problem (6), i.e. using affine (matrix) inequalities and an equality constraint. We obtain:

$$\begin{aligned} \min_{\eta, \mathcal{K}, \mathcal{P}_1, \mathcal{P}_2, \mathcal{Z}, \mathcal{E}_1, \mathcal{E}_2} \quad & \eta^2, \\ \text{s.t.} \quad & \mathcal{P}_1 \succ 0,\ \mathcal{P}_2 \succ 0, \\ & \begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix} = \begin{pmatrix} \mathcal{P}_1 \\ \mathcal{P}_2 \end{pmatrix} \mathcal{B}_u \mathcal{K}, \\ & \begin{pmatrix} \mathcal{P}_2 & (\mathcal{C}_2 + \mathcal{D}_2 \mathcal{K})^T \\ \star & \mathcal{Z} \end{pmatrix} \succ 0, \\ & \operatorname{trace}(\mathcal{Z}) < \eta^2, \end{aligned}$$

$$\begin{pmatrix} \operatorname{sym}(\mathcal{P}_1 \mathcal{A} + \mathcal{E}_1) & \mathcal{C}_1 + \mathcal{D}_1 \mathcal{K} & \mathcal{P}_1 \mathcal{B}_w \\ \star & -I & 0 \\ \star & \star & -\gamma^2 I \end{pmatrix} \prec 0,$$

$$\begin{pmatrix} \operatorname{sym}(\mathcal{P}_2 \mathcal{A} + \mathcal{E}_2) & \mathcal{P}_2 \mathcal{B}_w \\ \star & -I \end{pmatrix} \prec 0.$$

In Figure 1 the equality constraint violation $\|\mathcal{E} - \mathcal{P}\mathcal{B}_u\mathcal{K}\|_*$ is plotted for three different values of λ and for five initial values X_a^1, Y_a^1 (the same initial values for every λ and for both algorithms). From this figure it is clear that for $\lambda = 5$ no solutions to problem (1) are found with neither Algorithm 1

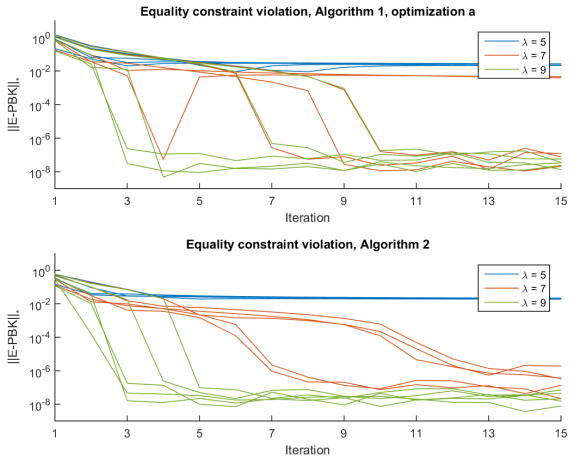

Fig. 1. The nuclear norm of the equality constraint violation, $\|\mathcal{E} - \mathcal{P}\mathcal{B}_u\mathcal{K}\|_*$.

nor 2 for all initial conditions. For $\lambda = 7$, Algorithm 1 finds a solution for some of the cases, and Algorithm 2 for all these initial conditions. For $\lambda = 9$ both algorithms successfully find solutions. What can also be observed is that Algorithm 2 is quicker to converge to a solution of problem (1) for, for example, $\lambda = 9$, even when we cannot provide a guarantee that it will. Finally, it can be seen that for Algorithm 1 a guarantee of convergence for $\eta^2 + \lambda \|\mathcal{E} - \mathcal{P}\mathcal{B}_u\mathcal{K}\|_*$, does not imply convergence of $\|\mathcal{E} - \mathcal{P}\mathcal{B}_u\mathcal{K}\|_*$ to 0,[1] since at iteration 4/5 one of the cases actually converges away from 0.

To obtain these results, we used the Mosek 7.1 [24] solver in conjunction with Matlab and Yalmip [20].

The best performance obtained with Algorithm 1 and 2 are $\mathcal{H}_2$ norms of 0.7507 and 0.7492 respectively, for which λ was set to a value of 6.85. The resulting controllers are

$$\mathcal{K}_{\text{alg 1}} = \begin{pmatrix} 1.8577 & 0.3003 & -0.3271 \end{pmatrix},$$
$$\mathcal{K}_{\text{alg 2}} = \begin{pmatrix} 1.9711 & 0.4222 & -0.1482 \end{pmatrix}.$$

For a state of the art method like implemented in PENBMI, from PENOPT [23] a solution was reported in [12] with an $\mathcal{H}_2$ norm of 0.74894, only a fraction lower than our methods. The method in [23] combines ideas of penalty methods and augmented Lagrangian methods and is a direct BMI solving programme, but the PENBMI does not use a rank constraint as we do and the methods described in [15], [14], but treats the BMI problem in a general context of augmented Lagrangian methods.

[1] We attribute the remaining difference of order 10^{-8} to the numerical settings of the SDP solver.

REFERENCES

[1] C. Scherer and S. Weiland, "Linear matrix inequalities in control," *Lecture Notes, Dutch Institute for Systems and Control, Delft, The Netherlands*, 2000.

[2] C. Scherer, P. Gahinet, and M. Chilali, "Multiobjective output-feedback control via lmi optimization," *Automatic Control, IEEE Transactions on*, vol. 42, no. 7, pp. 896–911, 1997.

[3] I. Masubuchi, A. Ohara, and N. Suda, "Lmi-based controller synthesis: A unified formulation and solution," *International Journal of Robust and Nonlinear Control*, vol. 8, no. 8, pp. 669–686, 1998.

[4] O. Toker and H. Özbay, "On the np-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback," in *American Control Conference, Proceedings of the 1995*, vol. 4, pp. 2525–2526, IEEE, 1995.

[5] S. Kanev, C. Scherer, M. Verhaegen, and B. De Schutter, "Robust output-feedback controller design via local bmi optimization," *Automatica*, vol. 40, no. 7, pp. 1115–1127, 2004.

[6] M. Fukuda and M. Kojima, "Branch-and-cut algorithms for the bilinear matrix inequality eigenvalue problem," *Computational Optimization and Applications*, vol. 19, no. 1, pp. 79–105, 2001.

[7] H. Tuan, P. Apkarian, S. Hosoe, and H. Tuy, "Dc optimization approach to robust control: feasibility problems," *International Journal of Control*, vol. 73, no. 2, pp. 89–104, 2000.

[8] T. Iwasaki, "The dual iteration for fixed-order control," *IEEE transactions on automatic control*, vol. 44, no. 4, pp. 783–788, 1999.

[9] K.-C. Goh, M. G. Safonov, and G. P. Papavassilopoulos, "Global optimization for the biaffine matrix inequality problem," *Journal of global optimization*, vol. 7, no. 4, pp. 365–380, 1995.

[10] Y. Yamada and S. Hara, "Global optimization for h_∞ control with constant diagonal scaling," *Automatic Control, IEEE Transactions on*, vol. 43, no. 2, pp. 191–203, 1998.

[11] A. Hassibi, J. How, and S. Boyd, "A path-following method for solving bmi problems in control," in *American Control Conference, 1999. Proceedings of the 1999*, vol. 2, pp. 1385–1389, IEEE, 1999.

[12] E. Ostertag, "An improved path-following method for mixed controller design," *Automatic Control, IEEE Transactions on*, vol. 53, no. 8, pp. 1967–1971, 2008.

[13] K. Goh, M. Safonov, and G. Papavassilopoulos, "A global optimization approach for the bmi problem," in *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, vol. 3, pp. 2009–2014, IEEE, 1994.

[14] S. Ibaraki and M. Tomizuka, "Rank minimization approach for solving bmi problems with random search," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 3, pp. 1870–1875, IEEE, 2001.

[15] S.-J. Kim and Y.-H. Moon, "Structurally constrained h_2 and h_∞ control: A rank-constrained lmi approach," *Automatica*, vol. 42, no. 9, pp. 1583–1588, 2006.

[16] D. Carlson, E. Haynsworth, and T. Markham, "A generalization of the schur complement by means of the moore-penrose inverse," *SIAM Journal on Applied Mathematics*, vol. 26, no. 1, pp. 169–175, 1974.

[17] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 9, pp. 2117–2130, 2013.

[18] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural computation*, vol. 15, no. 4, pp. 915–936, 2003.

[19] C. King, "Inequalities for trace norms of 2×2 block matrices," *Communications in mathematical physics*, vol. 242, no. 3, pp. 531–545, 2003.

[20] J. Löfberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pp. 284–289, IEEE, 2004.

[21] P. P. Khargonekar, M. Rotea, *et al.*, "Mixed h_2/h_∞ control: a convex optimization approach," *Automatic Control, IEEE Transactions on*, vol. 36, no. 7, pp. 824–837, 1991.

[22] J.-B. Thevenet, D. Noll, and P. Apkarian, "Non linear spectral sdp method for bmi-constrained problems: Applications to control design," in *Informatics in Control, Automation and Robotics I*, pp. 61–72, Springer, 2006.

[23] M. Kočvara and M. Stingl, "Pennon: A code for convex nonlinear and semidefinite programming," *Optimization methods and software*, vol. 18, no. 3, pp. 317–333, 2003.

[24] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2015.