Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 53-2 (2020) 5837–5842

An alternating projection algorithm for the "approximate" GCD calculation

Limantseva O. * **Halikias G.** ** **Karcanias N.** ***

* *School of Mathematics, Computer Science and Engineering, City, University of London, Northampton Square, London EC1V 0HB, UK (e-mail: olga.limantseva@city.ac.uk)*

** *School of Mathematics, Computer Science and Engineering, City, University of London, Northampton Square, London EC1V 0HB, UK (e-mail: G.Halikias@city.ac.uk)*

*** *School of Mathematics, Computer Science and Engineering, City, University of London, Northampton Square, London EC1V 0HB, UK (e-mail: N.Karcanias@city.ac.uk)*

Abstract: In the paper an approach is proposed for calculating the "best" approximate GCD of a set of coprime polynomials. The algorithm is motivated by the factorisation of the Sylvester resultant matrix of polynomial sets with nontrivial GCD. In the (generic) case of coprime polynomial sets considered here the aim is to minimise the norm of the residual error matrix of the inexact factorisation in order to compute the "optimal" approximate GCD. A least-squares alternating projection algorithm is proposed as an alternative to the solution of the corresponding optimisation problem via nonlinear programming techniques. The special structure of the problem in this case, however, means that the algorithm can be reduced to a sequence of standard subspace projections and hence no need arises to compute gradient vectors, Hessian matrices or optimal step-lengths. An estimate of the asymptotic convergence rate of the algorithm is finally established via the inclination of two subspaces.

Keywords: nonlinear least squares, Sylvester resultant matrix, approximate GCD, alternating projection algorithm, convergence analysis

1. INTRODUCTION

The notion of Greatest Common Divisors (GCD) of a set of polynomials has a wide range of applications in engineering disciplines. In Control Theory, for example, they arise in determinantal assignment problems, algebraic control methods, stability analysis problems, or distance problems to uncontrollability or unobservability.

Various approaches have been developed for GCD calculations, e.g. Barnett (1971), Bitmead et al. (1978), Fatouros and Karcanias (2003), Vardulakis and Stoyle (1978). Euclid's algorithm for two polynomials established the basis for the development of several computational techniques that have been extended to arbitrary number of polynomials. Linear Algebraic methods for GCD calculation rely on the properties of the corresponding Sylvester resultant matrix. Certain of these techniques have been extensively studied in Christou et al. (2010), Fatouros and Karcanias (2003), Karcanias and Mitrouli (2003), Karcanias and Mitrouli (1994), Karcanias et al. (2006b), Karcanias (2007), Karmarkar and Lakshman (1996), Mitrouli and Karcanias (1993), Rupprecht (1999), and the references therein.

The notion of "almost" or "approximate" coprimeness addressed in Mitrouli and Karcanias (1993), Karcanias and Mitrouli (1994), Fatouros et al. (2013), Halikias et al. (2012), Karcanias (2007), is of particular interest because the GCD computation is a non-generic problem. Algorithms for the approximate GCD calculation typically relax algorithms for solving the exact GCD problem and have found application in a variety of contexts, e.g. Karcanias et al. (2006a), Karcanias and Halikias (2013). Mitrouli and Karcanias (1993) propose an algorithm based on the ERES methodology, while Karcanias and Giannakopoulos (1989), consider matrix-pencil techniques for computation of the nearest GCD arbitrary degree. Bitmead et al. (1978), Fatouros and Karcanias (2003), Halikias et al. (2012), Karcanias et al. (2006a) characterise approximate GCD's based on the properties of Sylvester resultant matrices. An alternative approach has been studied in Halikias et al. (2012), Limantseva et al. (2020) based on the concept of structured singular values.

In the present work we propose an approach based on a factorisation result of the generalised Sylvester resultant matrix defined by a set of polynomials having GCD of degree k. It is well known that a set of polynomials $\mathcal{P}$ share a common root if and only if the corresponding Sylvester matrix, say $S_{\mathcal{P}}$, is rank deficient. If the GCD has degree k, then it can be written as a product of an augmented resultant matrix $[0_k \mid S_{\mathcal{P}^*}]$ corresponding to a set of polynomials $\mathcal{P}^*$ of reduced degree and a lower-triangular Toeplitz matrix $\hat{\Phi}_\lambda$ defined from the coefficients of the GCD polynomial, i.e. $S_{\mathcal{P}} = [0_k \mid S_{\mathcal{P}^*}]\hat{\Phi}_\lambda$. If $\mathcal{P}$ has a GCD with degree less than k, then the factorisation is not exact and

10.1016/j.ifacol.2020.12.1630

the residual error matrix $E = S_{\mathcal{P}} - [0_k \mid S_{\mathcal{P}^*}]\hat{\Phi}_\lambda$ occurs. The proposed algorithm aims to minimise the Frobenious norm of this matrix with respect to the parameters of the factorised matrices $S_{\mathcal{P}^*}$ and $\hat{\Phi}_\lambda$. This minimisation is equivalent to a nonlinear programming problem. Here we propose a least-squares alternating projection algorithm to minimise $\|E\|_F$ involving a sequence of linear projections. This has the advantage that it reduces the optimization to a sequence of standard subspace projections and hence no need arises to compute gradient vectors, Hessian matrices or optimal step-lengths. Estimates of the asymptotic convergence rate can also de derived in terms of the inclination of two subspaces.

The notation used in the paper is standard and is summarised in this paragraph for convenience. We denote by $\mathbb{R}^{n\times m}$ the set of $n \times m$ real matrices and by $\mathbb{R}[s]$ the set of polynomials with real coefficients. If $d(s) \in \mathbb{R}[s]$, then $\partial d(s)$ defines the degree of $d(s)$. The singular values of $A \in \mathbb{R}^{n\times m}$ are denoted by $\sigma_i(A)$, $i = 1, \ldots, \min(n, m)$, listed in non-increasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(n,m)} \geq 0$. $\|A\| = \sigma_1(A)$ denotes the spectral norm of a matrix A and $\|A\|_F$ the Frobenius norm. The generalised Moore-Penrose inverse of A is denoted by A^+. If P is the projection operator onto a subspace S of $\mathbb{R}^m$, then $P^\perp$ is the projection operator onto $S^\perp$, the orthogonal complement of S. Finally, the set of lower triangular Toepliz matrices is denoted by $\mathcal{T}$.

2. GENERALISED SYLVESTER RESULTANT AND GCD FACTORISATION

Let $\mathcal{P}_{h+1,n}$ be a set of polynomials defined as follows:

$$\mathcal{P}_{h+1,n} = \{a(s), b_i(s) \in \mathbb{R}[s],\ i = 1, \ldots, h\} \tag{1}$$

where $n = \partial a(s) \geq t = \partial b_i(s)$, $i = 1, 2, \ldots, h$. Let the GCD of the set $\mathcal{P}_{h+1,n}$ be $\phi(s)$ ($\phi(s) = 1$ if the polynomials are coprime). The crucial concept of studying properties of a polynomial set is the notion of Sylvester resultant matrix that is studied in Barnett (1971), Bitmead et al. (1978), Christou et al. (2010), Fatouros and Karcanias (2003), Karcanias et al. (2006b), Vardulakis and Stoyle (1978).

Definition 1. Let $\mathcal{P}_{h+1,n}$ as defined in (1) where $a(s)$ and $b_i(s)$, $\forall i = 1, \ldots, h$, are considered without loss of generality to be the monic polynomials:

$$\begin{aligned} a(s) &= s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_0, \\ b_i(s) &= s^t + \beta_{t-1,i}s^{t-1} + \cdots + \beta_{0,i},\ \ i = 1, \ldots, h \end{aligned} \tag{2}$$

(1) Let the Sylvester Resultant $S_0 \in \mathbb{R}^{t\times(n+t)}$ corresponding to $a(s)$ be structured as follows:

$$S_0 = \begin{pmatrix} 1 & \alpha_{n-1} & \alpha_{n-2} & \cdots & \cdots & \alpha_0 & 0 & \cdots & 0 \\ 0 & 1 & \alpha_{n-1} & \cdots & \cdots & \alpha_1 & \alpha_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \cdots & \alpha_{t-1} & \cdots & \cdots & \alpha_0 \end{pmatrix} \tag{3}$$

(2) For each polynomial $b_i(s)$, where $i = 1, \ldots, h$, define corresponding Resultant matrices by $S_i \in \mathbb{R}^{n\times(n+t)}$ where:

$$S_i = \begin{pmatrix} 1 & \beta_{i,t-1} & \beta_{i,t-2} & \cdots & \cdots & \beta_{i,0} & 0 & \cdots & 0 \\ 0 & 1 & \beta_{i,t-1} & \cdots & \cdots & \beta_{i,1} & \beta_{i,0} & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \cdots & \beta_{i,n-1} & \cdots & \cdots & \beta_{i,0} \end{pmatrix} \tag{4}$$

(3) Then, the Generalised Resultant of the polynomial set $\mathcal{P}_{h+1,n}$ is structured:

$$S_{\mathcal{P}} = \begin{pmatrix} S_0 \\ S_1 \\ \vdots \\ S_h \end{pmatrix} \in \mathbb{R}^{(t+hn)\times(n+t)} \tag{5}$$

The Generalised Resultant matrix holds important properties of the GCD of the given polynomial set. Moreover, the k-th smallest singular value of this matrix can be used as the approximate measure of the nearest GCD of degree k. Such concept is widely used in the matrix pencil methodologies, see Karcanias and Mitrouli (1994), Karcanias et al. (2006b), where the number of singular values within a given tolerance indicates the approximate GCD degree.

Many studies has been devoted to the problem of "approximate co-primness" and "almost zeros" of a set of polynomials, Christou et al. (2010), Fatouros et al. (2013), Karcanias et al. (2006b), Karcanias (2007), Karcanias et al. (2006a), Karcanias and Halikias (2013), Karmarkar and Lakshman (1996). In the context of our work, an almost rank-deficient Sylvester resultant matrix identifies the existence of an almost common root in the polynomial set, while the smallest singular value may be regarded as a measure of distance to non-coprimeness, Christou et al. (2010), Fatouros et al. (2013), Karcanias (2007).

Before introducing the main results some important properties of $S_{\mathcal{P}}$ are summarised in the following Theorem.

Theorem 1. Fatouros and Karcanias (2003): Consider the generalised Sylvester resultant $S_{\mathcal{P}}$ of the polynomial set $\mathcal{P}_{h+1,n}$ defined in equation (5).nLet $\phi(s)$ be the GCD of the polynomials in $\mathcal{P}_{h+1,n}$. Then:

(1) The polynomial set $\mathcal{P}_{h+1,n}$ is coprime if and only if the corresponding Sylvester resultant has full rank, i.e. $\text{rank}(S_{\mathcal{P}}) = n + t$. In general,

$$\text{rank}(S_{\mathcal{P}}) = n + t - \partial\phi(s) \tag{6}$$

where n, t are the maximal degrees of the polynomials $a(s)$ and $b_i(s)$ respectively, $i = 1, \ldots, h$.

(2) $\phi(s)$ is invariant under elementary row transformations on $S_{\mathcal{P}}$. The last non-vanishing row of the reduced row-echelon form of $S_{\mathcal{P}}$ defines the coefficients of $\phi(s)$.

Karcanias et al. (2006a) introduced technique for calculating the "best" approximate GCD based on the factorisation of the generalised resultant matrix. This technique forms the basis of the algorithm proposed in the paper.

Definition 2. Karcanias et al. (2006a): Let $\lambda(s) = \lambda_k s^k + \cdots + \lambda_1 s + \lambda_0$ be the GCD of the set $\mathcal{P}_{h+1,n}$. Assume that the Toeplitz matrix $\hat{\Phi}_\lambda \in \mathcal{T}$ corresponding to $\lambda(s)$ is structured as:

$$\hat{\Phi}_\lambda = \begin{pmatrix} \lambda_0 & 0 & 0 & \dots & \dots & \dots & 0 \\ \lambda_1 & \lambda_0 & 0 & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \lambda_k & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \lambda_k & & \lambda_1 & \lambda_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_k & \dots & \lambda_1 & \lambda_0 \end{pmatrix} \in \mathbb{R}^{(n+t)\times(n+t)}. \quad (7)$$

$\hat{\Phi}_\lambda$ is invertible if and only if $\lambda_0 \neq 0$. In that case the inverse, denoted as Φ_λ, is also a Toeplitz matrix representation, such that:

$$\Phi_\lambda = \begin{pmatrix} y_0 & 0 & 0 & \dots & \dots & 0 \\ y_1 & y_0 & 0 & & & \vdots \\ y_2 & y_1 & y_0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ y_{n+t-2} & y_{n+t-3} & \cdots & \cdots & y_0 & 0 \\ y_{n+t-1} & y_{n+t-2} & \cdots & \cdots & y_1 & y_0 \end{pmatrix} \in \mathbb{R}^{(n+t)\times(n+t)}, \quad (8)$$

where the elements y_i are

$$y_0 = \frac{1}{\lambda_0}, \;\; y_1 = \frac{\lambda_1}{\lambda_0} y_0, \dots, y_i = -\frac{1}{\lambda_0} \sum_{i=1}^{\min\{j,k\}} \lambda_i y_{j-i}, \quad (9)$$

and $j = 2, \dots, n+t-1$.

Theorem 2. Karcanias et al. (2006a): Let $S_\mathcal{P}$ be the Sylvester resultant matrix of the polynomial set $\mathcal{P}_{h+1,n}$. Set $\lambda(s) = \lambda_k s^k + \cdots + \lambda_1 s + \lambda_0$ be the greatest common divisor of degree k of the set. thus, it is possible to factor $S_\mathcal{P}$ as

$$S_\mathcal{P} = S^{(k)}_{\mathcal{P}^*} \hat{\Phi}_\lambda, \;\; S^{(k)}_{\mathcal{P}^*} = \begin{pmatrix} 0_k & S^{(k)}_0 \\ 0_k & S^{(k)}_1 \\ \vdots & \vdots \\ 0_k & S^{(k)}_h \end{pmatrix} = [0_k \mid \mathcal{S}_{p^*}]$$

where $\mathcal{S}^*_p$ is the reduced set of polynomials $\mathcal{P}^*_{h+1,n-k}$ obtained by dividing the polynomials of the set $\mathcal{P}_{h+1,n}$ by $\lambda(s)$,

$$S^{(k)}_0 = \begin{pmatrix} \alpha^{(k)}_{n-k} & \alpha^{(k)}_{n-k-1} & \dots & \alpha^{(k)}_0 & 0 & \dots & 0 \\ 0 & \alpha^{(k)}_{n-k} & \dots & \dots & \alpha^{(k)}_0 & \dots & 0 \\ \vdots & \ddots & \ddots & & & \ddots & \vdots \\ 0 & \dots & 0 & \alpha^{(k)}_{n-k} & \dots & \dots & a^{(k)}_0 \end{pmatrix} \quad (10)$$

and

$$S^{(k)}_i = \begin{pmatrix} \beta^{(k)}_{i,t-k} & \beta^{(k)}_{i,t-k-1} & \dots & \beta^{(k)}_{i,0} & 0 & \dots & 0 \\ 0 & \beta^{(k)}_{i,t-k} & \dots & \dots & \beta^{(k)}_{i,0} & \dots & 0 \\ \vdots & \ddots & \ddots & & & \ddots & \vdots \\ 0 & \dots & 0 & \beta^{(k)}_{i,t-k} & \dots & \dots & \beta^{(k)}_{i,0} \end{pmatrix} \quad (11)$$

for each $i = 1, \dots, h$. Moreover, $\hat{\Phi}_\lambda$ is the lower-triangular Toeplitz matrix defined in equation (7).

Equivalently, if $\lambda(s)$ does not have a zero root, then

$$S^{(k)}_{\mathcal{P}^*} = S_\mathcal{P} \Phi_\lambda = [0_k \mid S_{\mathcal{P}^*}]$$

where Φ_λ is the lower-triangular Toeplitz matrix defined in equations (8) and (9).

Note that factorisation of the generalised resultant is exact if and only if the polynomial set $\mathcal{P}_{h+1,n}$ has exact GCD. Otherwise the GCD is approximate and leads to the non-zero residual error matrix:

$$E = S_\mathcal{P} - S^{(k)}_{\mathcal{P}^*} \hat{\Phi}_\lambda = S_\mathcal{P} - [0_k \mid \mathcal{S}_{\mathcal{P}^*}] \hat{\Phi}_\lambda \quad (12)$$

which depends on the parameters:

$$\{\alpha^{(k)}_j\}_{j=0,1,\dots,n-k-1}, \;\; \{\beta^{(k)}_{i,j}\}^{j=0,1,\dots,t-k-1}_{i=1,2,\dots,h} \text{ and } \{\lambda_j\}_{j=0,1,\dots k-1} \quad (13)$$

To make the parametrization unique it is assumed that all polynomials are monic, i.e. $\lambda_k = 1$, $\alpha^{(k)}_{n-k} = 1$ and $\beta^{(k)}_{i,t-k} = 1$ for all $i = 1, 2, \dots, h$.

In the proposed algorithm, we seek to minimise the following function:

$$f(\theta) := \|S_\mathcal{P} - [0_k \mid S_{\mathcal{P}^*}] \hat{\Phi}_\lambda\|^2_F \quad (14)$$

with respect to parameter vector

$$(\theta, \lambda) = (\alpha^{(k)}_0, \dots \alpha^{(k)}_{n-k-1}, \beta^{(k)}_{1,0}, \dots \beta^{(k)}_{h,t-k-1}, \lambda_0, \dots \lambda_{k-1}) \quad (15)$$

whose elements are defined in equation (13). The optimal parameters obtained by the algorithm define the "best" approximate GCD and the corresponding coefficients of the reduced-degree polynomials, along with the minimum value of f which is related to the strength of the approximation.

Equation (14) is a nonlinear least squares-problem. Such problems have been of particular interest for many years, see Marquardt (1963), Powell and Macdonald (1972), Wold (1973) Golub and Pereyra (1973), Ruhe and Wedin (1980), O'leary and Rust (2013), Yamashita and Fukushima (2001), Bergou et al. (2017), Argyros et al. (2019). In the next section we analyse the convergence properties of the special family of separable least-squares problems which includes the specific problem under consideration in this work.

3. SEPARABLE PROBLEMS

In Ruhe and Wedin (1980), Golub and Pereyra (1973), Golub and Pereyra (2003) and Krogh (1974) the authors consider a special class of separable problems of the form:

$$\min\{\|y - \Psi(\theta)\lambda\|, \; \theta \in \mathbb{R}^n, \lambda \in \Omega\} \quad (16)$$

where Ω is an open subset of $\mathbb{R}^m$. This is a nested optimization problem over parameter vectors θ and λ. Since dependence on λ is linear, minimization over λ can be solved explicitly by means of a subspace projection (parametric in θ) resulting in:

$$\min_{\theta\in\Omega} \|P^\perp_{\Psi(\theta)} y\|$$

which is a nonlinear programming problem. This approach simplifies the computational procedure and appears to be more robust in comparison with methodologies based on the Taylor series expansion or gradient-based techniques (see Golub and Pereyra (2003), Krogh (1974)). Local and global optimality conditions in terms of the transformed problem can be found in Golub and Pereyra (1973).

The present problem falls in the above class. Consider the function:

$$f(\theta, \lambda) = \|S_\mathcal{P} - [0_k \mid S_{\mathcal{P}^*}] \hat{\Phi}_\lambda\|^2_F \quad (17)$$

where parameter vectors are $\lambda = [\lambda_0, \dots \lambda_{k-1}]'$ and $\theta = [\alpha_0^{(k)}, \dots \alpha_{n-k-1}^{(k)}, \beta_{1,0}^{(k)}, \dots \beta_{h,t-k-1}^{(k)}]'$. Since the θ_i's enter $S_{\mathcal{P}^*}$ linearly (and the λ_i's enter $\hat{\Phi}_\lambda$ linearly), the problem is separable. In particular we can write:

$$\begin{aligned} f(\theta,\lambda) &= \|S_{\mathcal{P}} - [0_k \mid S_{\mathcal{P}^*}]\hat{\Phi}_\lambda\|_F^2 \\ &= \|\mathrm{vec}(S_{\mathcal{P}}) - (I \otimes [0_k \mid S_{\mathcal{P}^*}])\mathrm{vec}(\hat{\Phi}_\lambda)\|^2 \end{aligned} \quad (18)$$

Let Q be such that $\mathrm{vec}(\hat{\Phi}_\lambda) = Q\lambda$, then

$$f(\theta,\lambda) = \|\mathrm{vec}(S_{\mathcal{P}}) - (I \otimes [0_k \mid S_{\mathcal{P}^*}])Q\lambda\|^2.$$

Identifying $y := \mathrm{vec}(S_{\mathcal{P}})$ and $\Psi(\theta) = (I \otimes [0_k \mid S_{\mathcal{P}^*}])Q$ shows that $f(\theta,\lambda)$ is in the form of (16). Note further that due to the bilinearity of the problem this can also be achieved by interchanging the role of the two parameter vectors. Writing:

$$\begin{aligned} f(\theta,\lambda) &= \|S_{\mathcal{P}} - [0_k \mid S_{\mathcal{P}^*}]\hat{\Phi}_\lambda\|_F^2 \\ &= \|\mathrm{vec}(S_{\mathcal{P}}) - (\hat{\Phi}_\lambda \otimes I)\mathrm{vec}([0_k \mid S_{\mathcal{P}^*}])\|^2 \end{aligned}$$

and introducing a matrix $\hat{Q}$ so that

$$\mathrm{vec}([0_k \mid S_{\mathcal{P}^*}]) = \hat{Q}\theta$$

then $f(\theta,\lambda)$ can again be written in the general form of (16).

Next we exploit the bilinearity of the cost function to introduce a least-squares alternating projection algorithm for minimising the residual error function (14).

4. AN ALTERNATING PROJECTION LEAST-SQUARES ALGORITHM FOR CALCULATING THE APPROXIMATE GCD

Let $\mathcal{P}_{h+1,n}$ be a set of monic polynomials with dominant degrees (n,t), $t \le n$, and let $S_{\mathcal{P}}$ be the corresponding generalised Sylvester resultant. Assume that there exists an approximate GCD of degree k defined by the monic polynomial $\lambda(s) = s^k + \lambda_{k-1}s^{k-1} + \cdots + \lambda_1 s + \lambda_0$, $k \le t$. Equation (7) defines the Toeplitz matrix of the GCD, where λ_k is assumed to be equal to 1. In order to find "best" set of GCD coefficients and optimise the strength of approximation we define $E = S_{\mathcal{P}} - [0_k \mid S_{\mathcal{P}^*}]\hat{\Phi}_\lambda$ and seek to minimise $f = \|E\|_F^2$. Then the problem can be written as follows:

$$\begin{aligned} E = S_{\mathcal{P}} - \left(\tilde{N}_0 + \sum_{j=0}^{n-k-1} \alpha_j^{(k)} N_j + \sum_{i=1}^{h}\sum_{j=0}^{t-k-1} \beta_{ij}^{(k)} N_{ij}\right) \times \\ \times \left(\tilde{M}_0 + \sum_{i=0}^{k-1} \lambda_i M_i\right) \end{aligned} \quad (19)$$

where $\tilde{N}_0, N_j, N_{ij} \in \mathbb{R}^{(t+hn)\times(n+t)}$, $\tilde{M}_0, M_i \in \mathbb{R}^{(n+t)\times(n+t)}$. The algorithms considers two separate subproblems that are solved iteratively in the following steps:

Step 1:

Consider the parameters $(\lambda_0, \lambda_1, \dots, \lambda_{k-1})$ to be fixed at the first step. Let

$$M = \tilde{M}_0 + \sum_{i=0}^{k-1} \lambda_i M_i = \hat{\Phi}_\lambda$$

Then the problem is reduced to:

$$\begin{aligned} E &= S_{\mathcal{P}} - \left(\tilde{N}_0 + \sum_{j=0}^{n-k-1} \alpha_j^{(k)} N_j + \sum_{i=1}^{h}\sum_{j=0}^{t-k-1} \beta_{ij}^{(k)} N_{ij}\right) M \\ &= \tilde{Y}_0 - \sum_{j=0}^{n-k-1} \alpha_j^{(k)} Y_j - \sum_{i=1}^{h}\sum_{j=0}^{t-k-1} \beta_{ij}^{(k)} Y_{ij} \end{aligned}$$

where $\tilde{Y}_0 = S_{\mathcal{P}} - \tilde{N}_0 M$, $Y_j = N_j M$ and $Y_{ij} = N_{ij} M$ for $i = 1, 2, \dots, h$ and $j = 0, 1, \dots, t-k-1$. We seek to minimise:

$$f_1(\alpha_0^{(k)}, \dots, \beta_{h,t-k-1}^{(k)}) = f_1(\theta), \quad (20)$$

where θ is defined in equation (15). This is equivalent to the minimisation of function:

$$f_1 = \left\| \tilde{Y}_0 - \sum_{j=0}^{n-k-1} \alpha_j^{(k)} Y_j - \sum_{i=1}^{h}\sum_{j=0}^{t-k-1} \beta_{ij}^{(k)} Y_{ij} \right\|_F^2 \quad (21)$$

This can be written as $c_1 - 2y'\theta + \theta'\Gamma\theta$ with $c_1 = \|\tilde{Y}_0\|_F^2$, $\theta' = \begin{pmatrix} \alpha_0^{(k)} & \dots & \alpha_{n-k-1}^{(k)} & \beta_{1,0}^{(k)} & \dots & \beta_{h,t-k-1}^{(k)} \end{pmatrix} \in \mathbb{R}^{n-k+h(t-k)}$ and

$$y' = \begin{pmatrix} y_0' & y_1' & \dots & y_h' \end{pmatrix} \in \mathbb{R}^{n-k+h(t-k)}$$

where

$$y_0' = \begin{pmatrix} \mathrm{trace}(\tilde{Y}_0 Y_0') & \dots & \mathrm{trace}(\tilde{Y}_0 Y_{n-k-1}') \end{pmatrix} \in \mathbb{R}^{n-k}$$

and

$$y_i' = \begin{pmatrix} \mathrm{trace}(\tilde{Y}_0 Y_{i,0}') & \dots & \mathrm{trace}(\tilde{Y}_0 Y_{i,t-k-1}') \end{pmatrix} \in \mathbb{R}^{t-k}$$

for $i = 1, 2, \dots, h$. Moreover

$$\Gamma = \Gamma' = \begin{pmatrix} \Gamma^{0,0} & \Gamma^{0,1} & \dots & \Gamma^{0,h} \\ \Gamma^{1,0} & \Gamma^{1,1} & \dots & \Gamma^{1,h} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma^{h,0} & \Gamma^{h,1} & \dots & \Gamma^{h,h} \end{pmatrix},$$

$\Gamma^{0,0}_{i_0,j_0} = \mathrm{trace}(Y_{i_0-1}Y_{j_0-1}'), \Gamma^{0,m}_{i_0,j} = \mathrm{trace}(Y_{i_0-1}Y_{m,j-1}')$, $\Gamma^{\rho,m}_{i,j} = \mathrm{trace}(Y_{\rho,i-1}Y_{m,j-1}')$ and $i_0 = 1, 2, \dots, n-k$, $j_0 = 1, 2, \dots, n-k$, $i, j = 1, 2, \dots, t-k$, $m, \rho = 1, 2, \dots, h$; and $\Gamma \in \mathbb{R}^{(n-k-h(t-k))\times(n-k-h(t-k))}$. Since Γ is positive definite (see Lemma 1 below), the (unique) minimiser of f_1 is found as $\hat{\theta} = \Gamma^{-1}y$.

Step 2:

Assume that $\{\alpha_j^{(k)}\}_{j=0,1,\dots,n-k-1}$ and $\{\beta_{i,j}^{(k)}\}_{i=1,2,\dots,h}^{j=0,1,\dots,t-k-1}$ are fixed. Define:

$$N = [0_k \mid S_{\mathcal{P}}^*] = \tilde{N}_0 + \sum_{j=0}^{n-k-1} \alpha_j^{(k)} N_j + \sum_{i=1}^{h}\sum_{j=0}^{t-k-1} \beta_{ij}^{(k)} N_{ij}$$

Then the residual error matrix is:

$$E = S_{\mathcal{P}} - (N\tilde{M}_0 + \sum_{i=0}^{k-1} \lambda_i N M_i) = \tilde{X}_0 - \sum_{i=0}^{k-1} \lambda_i X_i$$

where $\tilde{X}_0 = S_{\mathcal{P}} - N\tilde{M}_0$ and $X_i = NM_i$, $i = 0, 1, \dots, k-1$. The second least-squares subproblem is to minimise the function:

$$f_2(\lambda_0, \dots, \lambda_{k-1}) = f_2(\lambda) \quad (22)$$

where θ is defined in equation (15). This is equivalent to the minimisation of:

$$\left\| \tilde{X}_0 - \sum_{i=0}^{k-1} \lambda_i X_i \right\|_F^2 = c - 2\sum_{i=0}^{k-1} \eta_i \lambda_i + \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \sigma_{ij}\lambda_i\lambda_j \quad (23)$$

or equivalently the minimization of: $f_2 = c_2 - 2\eta'\lambda + \lambda'\Sigma\lambda$, where we have defined:

$$c_2 = \|\tilde{X}_0\|_F^2, \;\; \lambda' = (\, \lambda_0 \;\; \lambda_1 \;\; \cdots \;\; \lambda_{k-1} \,) \in \mathbb{R}^k$$

and also

$$\eta' = (\, \eta_0 \;\; \eta_1 \;\; \cdots \;\; \eta_{k-1} \,) \in \mathbb{R}^k, \;\; \eta_i = \text{trace}(\tilde{X}_0 X_i')$$

for $i = 0, 1, \ldots, k-1$ and

$$\Sigma = \Sigma' \in \mathbb{R}^{k\times k} \;\text{ where }\; \Sigma_{ij} = \text{trace}(X_i X_j')$$

for $i = 0, 1, \ldots, k-1$ and $j = 0, 1, \ldots, k-1$. According to Lemma 1 (below) $\Sigma = \Sigma' > 0$ and hence the (unique) minimiser of f_2 is given by:

$$\hat{\lambda} = \left(\, \hat{\lambda}_0 \;\; \hat{\lambda}_1 \;\; \cdots \;\; \hat{\lambda}_{k-1} \,\right)' = \Sigma^{-1}\eta$$

Step 3:

Starting from an initial vector:

$$\hat{\lambda}' = \left(\, \hat{\lambda}_0 \;\; \hat{\lambda}_1 \;\; \cdots \;\; \hat{\lambda}_{k-1} \,\right)$$

whose entries are the initial estimates of the approximate GCD iterate between steps 2 and 3 of the algorithm until numerical convergence is attained, i.e. the difference between two consecutive values of f falls below a pre-specified tolerance.

The positive definiteness of matrices Γ and Σ above is established in the following Lemma.

Lemma 1. $\Gamma = \Gamma' > 0$ and $\Sigma = \Sigma' > 0$.

Proof: For $A, B \in \mathbb{R}^{(t+hn)\times(n+t)}$ define the inner product $(A, B) = \text{trace}(AB')$ and note that $(A, A) = \|A\|_F^2$. It can be shown that Γ is the Gramian matrix of the set

$$\{Y_0, \ldots, Y_{n-k-1}, Y_{1,0}, \ldots Y_{1,t-k-1}, \ldots, Y_{h,0}, \ldots Y_{h,t-k-1}\}$$

and that this set is linearly independent which implies $\Gamma = \Gamma' > 0$. A similar argument shows that $\Sigma = \Sigma' > 0$. Details are omitted. □

5. RATE OF CONVERGENCE OF ALTERNATING PROJECTION ALGORITHM

In this section the asymptotic rate of convergence of the alternation projection algorithm is studied. To simplify notation define:

$$N_{n+(i-1)t-ik+j} = N_{i,j}, \;\; 1 \le i \le h, \; 0 \le j \le t-k-1$$

and $\theta_i = \alpha_{i-1}^{(k)}$ for $i = 1, 2, \ldots, n-k-1$, $\theta_{n+(i-1)t-ik+j} = \beta_{i,j}^{(k)}$, $1 \le i \le h$, $0 \le j \le t-k-1$. Then

$$E = S_{\mathcal{P}} - \left(\tilde{N}_0 + \sum_{i=0}^{p}\theta_i N_i\right)\left(\tilde{M}_0 + \sum_{i=0}^{q}\lambda_i M_i\right)$$

where $p = n-k-1$ and $q = k-1$. Suppose that the alternating projection algorithm converges to the optimal solution with parameters $\{\theta_i^o\}_{i=0}^p$, $\{\lambda_j^o\}_{j=0}^q$ and consider small perturbations $\{\delta\theta_i\}_{i=0}^p$, $\{\delta\lambda_j\}_{j=0}^q$ around the optimal set of parameters. Then,

$$E = S_{\mathcal{P}} - \left(\tilde{N}_0 + \sum_{i=0}^{r}(\theta_i^o + \delta\theta_i)N_i\right)\left(\tilde{M}_0 + \sum_{i=0}^{q}(\lambda_i^o + \delta_i^o)M_i\right)$$

This is equivalent to

$$\begin{aligned} E = Y - \sum_{i=0}^{p}\delta\theta_i N_i(\tilde{M}_0 + \sum_{j=0}^{q}\lambda_j^0 M_j) - \\ - \sum_{i=0}^{q}\delta\lambda_i(\tilde{N}_0 + \sum_{j=0}^{p}\theta_j^0 N_j)M_i - \sum_{i-0}^{p}\sum_{j=0}^{q}\delta\theta_i\delta\lambda_j N_i M_j \end{aligned} \tag{24}$$

where

$$\begin{aligned} Y = S_{\mathcal{P}} - \tilde{N}_0\tilde{M}_0 - \sum_{i=0}^{p}\theta_i^o N_i\tilde{M}_0 - \sum_{i=0}^{q}\lambda_i^o\tilde{N}_0 M_i - \\ - \sum_{i-0}^{p}\sum_{j=0}^{q}\delta\theta_i^o\delta\lambda_j^o N_i M_j \end{aligned}$$

Vectorising E it can be observed that as $\delta\theta_i \to 0$ and $\delta\lambda_i \to 0$, $\|E\|_F^2 = \|\text{vec}(E)\|^2$ gets arbitrarily close to

$$\gamma = \|y - S_1\delta\theta - S_2\delta\lambda\|^2 \tag{25}$$

where $y = \text{vec}(Y)$,

$$(\delta\theta)' = (\, \delta\theta_1 \;\; \delta\theta_2 \;\; \ldots \;\; \delta\theta_p \,), \;\; (\delta\lambda)' = (\, \delta\lambda_1 \;\; \delta\lambda_2 \;\; \ldots \;\; \delta\lambda_q \,)$$

and

$$S_1 = (\, \xi_1 \;\; \xi_2 \;\; \ldots \;\; \xi_p \,), \;\; \xi_i = \text{vec}[N_i(\tilde{M}_0 + \sum_{j=0}^{q}\lambda_j^0 M_j)]$$

$$S_2 = (\, \psi_1 \;\; \psi_2 \;\; \ldots \;\; \psi_q \,), \;\; \psi_i = \text{vec}[(\tilde{N}_0 + \sum_{j=0}^{p}\theta_j^0 N_j)M_i]$$

where $i = 1, 2, \ldots, q$. Geometrically (25) corresponds to the distance of y from the subspace $\mathcal{R}(S_1) + \mathcal{R}(S_2)$, where $\mathcal{R}(S)$ is the range of subspace S. The following Theorem establishes the asymptotic convergence rate of the algorithm:

Theorem 3. (von Neumann) Let $\mathbb{X}$ be a Hilbert space and $\mathbb{U}$, $\mathbb{V}$, two closed subspaces of $\mathbb{X}$. Let also $\mathbb{W} = \overline{\mathbb{U} + \mathbb{V}}$ where $\overline{(\cdot)}$ denotes set closure. For $x \in \mathbb{X}$ define the sequence:

$$x_n = [(I - P_v)(I - P_u)]^n x, \; n \in \mathbb{N}$$

where P_v and P_u are the orthogonal projection operators onto $\mathbb{V}$ and $\mathbb{U}$, respectively. Then the sequence $\{x_n\}$, $n \in \mathbb{N}$ converges to $x - w$ where w is the best approximation to x from $\mathbb{W}$. Further convergence is geometric with a convergence rate:

$$\beta = \text{incl}(\mathbb{U}^\perp, \mathbb{V}^\perp)$$

where $\text{incl}(\cdot, \cdot)$ denotes the inclination between two subspaces, i.e.

$$\begin{aligned} \text{incl}(\mathbb{U}^\perp, \mathbb{V}^\perp) = \sup\{|(z, w)| : z \in \mathbb{U}^\perp, \; w \in \mathbb{V}^\perp, \\ \|z\| = \|w\| = 1\} \end{aligned}$$

and where $\mathbb{U}^\perp$ and $\mathbb{V}^\perp$ denote the orthogonal complements of $\mathbb{U}$ and $\mathbb{V}$, respectively.

The asymptotic convergence rate of the algorithm is obtained in the following Theorem (proof is omitted). This is based on von Neumann's algorithm. Note that since $\mathcal{R}(S_1)$ and $\mathcal{R}(S_2)$ are finite-dimensional they are closed (and similarly for $\mathcal{R}(S_1) + \mathcal{R}(S_2)$).

Theorem 4. Let U_2 and $\tilde{U}_2$ be matrices whose columns are orthonormal and span the range of $S_1^\perp$ and $S_2^\perp$, respectively. Then the asymptotic convergence rate of the algorithm is $\beta = \text{incl}(\mathcal{R}(S_1^\perp), \mathcal{R}(S_2^\perp)) = \sigma_1(U_2'\tilde{U}_2)$.

6. CONCLUSION

An alternative method for calculating the distance to the nearest common root for a set of polynomials is presented. The approximate factorisation of the Sylvester matrix motivated the definition of an alternating least-squares projection algorithm that minimises Frobenious norm of the error matrix of the factorisation. It is shown that a general nonlinear problem can be divided into two linear sub-problems and be solved iteratively. Such an approach avoids significant complexity in numerical calculations of the "best" approximate GCD of a coprime set of polynomials.

REFERENCES

I. Argyros, S. Shakhno, and Y. Shunkin. Improved convergence analysis of gauss-newton-secant method for solving nonlinear least squares problems. *Mathematics*, 7(1):99, 2019.

S. Barnett. Greatest common divisor of several polynomials. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 70, pages 263–268. Cambridge University Press, 1971.

E. Bergou, Y. Diouane, and V. Kungurtsev. Global and local convergence of a levenberg-marquadt algorithm for inverse problems. Technical report, Technical Report ISAE-SUPAERO, 2017.

R. Bitmead, S-Y. Kung, B. Anderson, and T. Kailath. Greatest common divisor via generalized sylvester and bezout matrices. *IEEE Transactions on Automatic Control*, 23(6):1043–1047, 1978.

D. Christou, N. Karcanias, and M. Mitrouli. The ERES method for computing the approximate GCD of several polynomials. *Applied Numerical Mathematics*, 60(1-2): 94–114, 2010.

S. Fatouros and N. Karcanias. Resultant properties of GCD of many polynomials and a factorization representation of GCD. *International Journal of Control*, 76 (16):1666–1683, 2003.

S. Fatouros, N. Karcanias, D. Christou, and P. Papadopoulos. Approximate greatest common divisor of many polynomials and pseudo-spectrum. *IFAC Proceedings Volumes*, 46(2):623–628, 2013.

G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems*, 19(2):R1, 2003.

G.H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis*, 10(2):413–432, 1973.

G. Halikias, G. Galanis, N. Karcanias, and E. Milonidis. Nearest common root of polynomials, approximate greatest common divisor and the structured singular value. *IMA Journal of Mathematical Control and Information*, 30(4):423–442, 2012.

N. Karcanias. The greatest common divisor of a set of polynomials, control theory and approximate algebraic computations. *BULLETIN OF THE GREEK MATHEMATICAL SOCIETY*, 54:29–46, 2007.

N. Karcanias and C. Giannakopoulos. Necessary and sufficient conditions for zero assignment by constant squaring down. *Linear Algebra and its applications*, 122: 415–446, 1989.

N. Karcanias and G. Halikias. Approximate zero polynomials of polynomial matrices and linear systems. *Linear Algebra and its Applications*, 439(4):1091–1103, 2013.

N. Karcanias and M. Mitrouli. A matrix pencil based numerical method for the computation of the GCD of polynomials. *IEEE Transactions on Automatic Control*, 39(5):977–981, 1994.

N. Karcanias and M. Mitrouli. Normal factorisation of polynomials and computational issues. *Computers & Mathematics with Applications*, 45(1-3):229–245, 2003.

N. Karcanias, S. Fatouros, M. Mitrouli, and G. Halikias. Approximate greatest common divisor of many polynomials, generalised resultants, and strength of approximation. *Computers & Mathematics with Applications*, 51(12):1817–1830, 2006a.

N. Karcanias, M. Mitrouli, and D. Triantafyllou. Matrix pencil methodologies for computing the greatest common divisor of polynomials: hybrid algorithms and their performance. *International Journal of Control*, 79(11): 1447–1461, 2006b.

N. Karmarkar and Y.N. Lakshman. Approximate polynomial greatest common divisors and nearest singular polynomials. In *Proceedings of the 1996 international symposium on Symbolic and algebraic computation*, pages 35–39. ACM, 1996.

F.T. Krogh. Efficient implementation of a variable projection algorithm for nonlinear least squares problems. *Communications of the ACM*, 17(3):167–169, 1974.

O. Limantseva, G. Halikias, and N. Karcanias. Nearest common root of a set of polynomials: A structured singular value approach. *Linear Algebra and its Applications*, 584(1):233–256, 2020.

D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

M. Mitrouli and N. Karcanias. Computation of the GCD of polynomials using gaussian transformations and shifting. *International Journal of Control*, 58(1):211–228, 1993.

D.P. O'leary and B.W. Rust. Variable projection for nonlinear least squares problems. *Computational Optimization and Applications*, 54(3):579–593, 2013.

D.R. Powell and J.R. Macdonald. A rapidly convergent iterative method for the solution of the generalised nonlinear least squares problem. *The Computer Journal*, 15(2):148–155, 1972.

A. Ruhe and P. Å. Wedin. Algorithms for separable nonlinear least squares problems. *SIAM review*, 22(3): 318–337, 1980.

D. Rupprecht. An algorithm for computing certified approximate GCD of n univariate polynomials. *Journal of Pure and Applied Algebra*, 139(1-3):255–284, 1999.

A.I. Vardulakis and P.N. Stoyle. Generalized resultant theorem. *IMA Journal of Applied Mathematics*, 22(3): 331–335, 1978.

H. Wold. Nonlinear iterative partial least squares (nipals) modelling: some current developments. In *Multivariate Analysis–III*, pages 383–407. Elsevier, 1973.

N. Yamashita and M. Fukushima. On the rate of convergence of the levenberg-marquardt method. In *Topics in numerical analysis*, pages 239–249. Springer, 2001.