Sparse Multivariate Function Recovery With a High Error Rate in the Evaluations*

Erich L. Kaltofen
Department of Mathematics
North Carolina State University
Raleigh, NC 27695-8205, USA
kaltofen@math.ncsu.edu www.kaltofen.us

Zhengfeng Yang
Shanghai Key Laboratory of Trustworthy Computing
East China Normal University
Shanghai 200062, China
zfyang@sei.ecnu.edu.cn

ABSTRACT

In [Kaltofen and Yang, Proc. ISSAC 2013] we have generalized algebraic error-correcting decoding to multivariate sparse rational function interpolation from evaluations that can be numerically inaccurate and where several evaluations can have severe errors ("outliers"). Here we present a different algorithm that can interpolate a sparse multivariate rational function from evaluations where the error rate is $1/q$ for any $q > 2$, which our ISSAC 2013 algorithm could not handle. When implemented as a numerical algorithm we can, for instance, reconstruct a fraction of trinomials of degree 15 in 50 variables with non-outlier evaluations of relative noise as large as 10^{-7} and where as much as 1/4 of the 14717 evaluations are outliers with relative error as small as 0.01 (large outliers are easily located by our method).

For the algorithm with exact arithmetic and exact values at non-erroneous points, we provide a proof that for random evaluations one can avoid quadratic oversampling. Our argument already applies to our original 2007 sparse rational function interpolation algorithm [Kaltofen, Yang and Zhi, Proc. SNC 2007], where we have experimentally observed that for T unknown non-zero coefficients in a sparse candidate ansatz one only needs $T+O(1)$ evaluations rather than the proven $O(T^2)$ (cf. Candès and Tao sparse sensing). Here we finally can give the probabilistic analysis for this fact.

Categories and Subject Descriptors: I.1.2 [Symbolic and Algebraic Manipulation]: Algorithms; G.1.1 [Numerical Analysis]: Interpolation—smoothing

Keywords: error correcting coding, fault tolerance, Cauchy interpolation, rational function

1. INTRODUCTION

Algorithms that interpolate a uni- or multivariate polynomial or rational function that reduce the number of values required for reconstruction according to the sparsity of polynomial in one of several bases, such as power basis, Chebychev basis, or Pochhammer basis, can be adapted to numerical data and floating point arithmetic. Error-correcting decoding of algebraic codes adds a new dimension to the domain of specifications: several evaluations can have erroneous values, which in the numeric setting constitute outliers with severe relative error. We present new exact and numeric algorithms for sparse multivariate rational function recovery when those outliers can be numerous, as much as 49.99% of all evaluations.

The literature on exact symbolic algorithms for sparse interpolation without errors is extensive. The first univariate algorithm is embedded in Blahut's [3] decoder for Reed/Solomon codes and generalized to sparse multivariate polynomials in [2]. A separate method for sparse multivariate polynomials is in [25, 24]. More efficient algorithms are in [12]. Early termination by randomization is introduced in [13]. Algorithms for sparse multivariate rational functions are in [7, 18, 6]. Finally, large degrees are dealt with in [8, 11, 14, 1]. Univariate algorithms for imprecise inputs go back to French revolution times [21], but came to life with early termination [9]. Low degree multivariate models can be recovered variable-by-variable [18]. Errors in the evaluations were introduced for univariate sparse polynomial recovery in [5], and for multivariate sparse polynomials and rational functions in [22, 17], the latter also with approximate evaluations and with outliers.

Our new algorithm adapts the variant of Zippel's [25] variable-by-variable iteration in [10, Section 4] as it has been applied to sparse multivariate rational function interpolation [18]. Those algorithms set up a large linear system with original values as entries (see (18)) and thus can tolerate imprecision in the entries, e.g., by use of a structured total least norm linear solver, as already observed in [18]. We deploy rational function interpolation as a means to decode errors in analogy to the dense univariate Olshevsky-Shokrollahi [20] decoder and unlike the Euclidean algorithm-based Berlekamp-Welch decoder [23, 15].

We shall not assume that the recovered rational fraction f/g is reduced, that is, we do **not** assume $\mathrm{GCD}(f, g) = 1$. The reasons are threefold: 1. sparse fractions can have fewer terms than reduced dense fractions, as in $(x^{20} - y^{20})/(x - y)$. Already in [14] we have interpolated such sparse unreduced fractions. 2. errors in evaluations at points are corrected by recovering $(f\Lambda)/(g\Lambda)$ where Λ is an error locator polynomial (see text above (4)). 3. approximate relative primeness is difficult to maintain numerically for many evaluations, as has already been observed in [16]. Avoiding the assumption because of 1. yields stability when the algorithm uses floating point arithmetic.

*This research was supported in part by the National Science Foundation under Grant CCF-1115772 (Kaltofen).

ISSAC'14, July 23–25, 2014, Kobe, Japan.

As in [17, 15], we allow evaluation of the rational function f/g at a pole, which we define as a zero of the (possibly unreduced) denominator g. We suppose that the evaluation at a pole is indicated by the resulting function value ∞. A true pole has that evaluation, but we also can have erroneous "false" poles, where ∞ is incorrectly indicated at a non-pole, or we can have an incorrect non-∞ value at a pole. The latter error can be located, but not corrected when the pole is simultaneously a root of the numerator f (see text below (17)).

Next we establish our notion of an error rate in a list of evaluations.

Remark 1.1. As in [17, Remark 1.1], we assume that the black box for evaluating f/g returns on a batch of $L \geq L_{\min}$ evaluations at distinct points no more than $k \leq L/q$ faulty values. Our model can be used to derive expected decoding results for a stochastic error rate via Chernoff bounds of the tail of the binomial distribution. If each evaluation is faulty with probability $1/\hat{q}$ then $\geq (1+\epsilon)L/\hat{q}$ evaluations are faulty with probability $\leq \exp(-\epsilon^2 L/(3\hat{q}))$ for all ϵ, $0 < \epsilon < 1$, and we may choose $q = \hat{q}/(1+\epsilon)$ in our analysis for an error probability of $1/\hat{q}$ with which a single evaluation is faulty. Therefore, from now on we shall refer to $1/q$ as the *adjusted* error rate. See also Lemma 3.1 below.

Our previous algorithm did not work for an adjusted error rate $1/q$ below but near to $1/2$ [17, Remark 2.6]. Our new algorithm achieves decoding with an error rate $1/q$ for any $q > 2$ by exploiting the dense univariate algorithm. However, we use the univariate algorithm only for error location, and derive the multivariate sparse rational function model f/g from a large homogeneous linear system (see (18)). The latter is necessary for imprecise data so that noise cannot be amplified in intermediately computed scalars. The size of that system is prescribed by the number of evaluations at random points that yield a proper interpolant Φ/Ψ in the sense that $f\Psi = \Phi g$, which we call the Cauchy property of the interpolating fraction Φ/Ψ.

Our algorithm restricts the polynomials Φ and Ψ to a sparsity structure which is overdetermined in a single variable and solves for the coefficients of Φ and Ψ from the values of $(f/g)(\xi_{1,\ell}, \ldots, \xi_{n,\ell})$. If we have T terms in the (overdetermined) combined term support set for the numerator Φ and denominator Ψ, we shall prove below that when the variable values $\xi_{\mu,\ell}$ are randomly and uniformly selected from a sufficiently large finite set of coefficient field elements, $T-1$ evaluation points yield the Cauchy property for all interpolants with high probability (see Theorem 2.1). Note that all scalar multiples $(c\,\Phi)/(c\,\Psi)$ are interpolants. The proof technique is a combination of row subsetting for solving overdetermined linear systems [19] and evaluation at new sets of symbols (transcendentals) [10, Section 4]. Earlier [18, 17], we only could provably establish the Cauchy property with $O(T^2)$ evaluations, based on the number of terms in $f\Psi - \Phi g$. We note the comparison to [4] where oversampling is quadratic in the sparsity.

Sections 2 and 3 analyze our algorithms when executed with exact arithmetic in the field of scalars. In Section 2 we describe an algorithm for recovering a vector of multivariate rational functions $[\,f^{\langle 1\rangle}/g, \ldots, f^{\langle s\rangle}/g\,]$. In Section 3 we describe a different algorithm that corrects at high error rate, using the analysis in Section 2 for a single rational function and without erroneous evaluations. The algorithm in Section 3 is easily generalized to vector recovery. Both sections must overcome the technical challenge of the asymmetric behavior of evaluations at poles that are simultaneously zeros of the numerator f (see (20) and the sentence thereafter). Section 4 describes the implementation and experiments with noisy data and floating point arithmetic.

2. VECTOR-OF-FUNCTIONS RECOVERY

The objective of this section is to generalize Kaltofen's upper bound estimate on the number of necessary function evaluations at random points [17, *Note added to Remark 2.2* posted on Kaltofen's web site version on July 14, 2013] to interpolating *a vector* of multivariate sparse rational functions with a common denominator:

$$[f^{\langle 1\rangle}/g, \ldots, f^{\langle s\rangle}/g] \in \mathsf{K}(x_1, \ldots, x_n)^s, \quad g \neq 0. \tag{1}$$

Note that the fractions $f^{\langle\sigma\rangle}/g$ are not necessarily reduced, and that may even have $\mathrm{GCD}(g, \mathrm{GCD}_\sigma(f^{\langle\sigma\rangle})) \neq 1$. We assume that we have for all σ, $1 \leq \sigma \leq s$, sets of terms $D_f^{\langle\sigma\rangle} \supseteq \mathrm{supp}(f^{\langle\sigma\rangle})$ that constitute maximal sparse supports, and a maximal sparse support set $D_g \supseteq \mathrm{supp}(g)$, where $\mathrm{supp}(f^{\langle\sigma\rangle})$ and $\mathrm{supp}(g)$ are the sets of terms in $f^{\langle\sigma\rangle}$ and g, resp., occurring with non-zero coefficients. As in [17], we evaluate the vector (1) ("probe the black box") at values for the variables, $(x_1, \ldots, x_n) \leftarrow (\xi_{1,\ell}, \ldots, \xi_{n,\ell}) \in \mathsf{K}^n$, for all L evaluations $0 \leq \ell \leq L-1$, where the $\xi_{\mu,\ell}$ are chosen in a certain way, e.g., selected randomly and uniformly from a finite subset $S \subseteq \mathsf{K}$. The obtained vector $[\,\beta_\ell^{\langle 1\rangle}, \ldots, \beta_\ell^{\langle s\rangle}] \in (\mathsf{K} \cup \{\infty\})^s$ can be incorrect in one or more components for $k \leq E$ evaluations $\ell = \lambda_1, \ldots, \lambda_k$, that is

$$\forall\kappa, 1\leq\kappa\leq k\colon \exists\sigma, 1\leq\sigma\leq s\colon \frac{f^{\langle\sigma\rangle}}{g}(\xi_{1,\lambda_\kappa}, \ldots, \xi_{n,\lambda_\kappa}) \neq \beta_{\lambda_\kappa}^{\langle\sigma\rangle}, \tag{2}$$

$$\forall\ell \notin \{\lambda_1, \ldots, \lambda_k\}\colon \forall\sigma, 1\leq\sigma\leq s\colon \frac{f^{\langle\sigma\rangle}}{g}(\xi_{1,\ell}, \ldots, \xi_{n,\ell}) = \beta_\ell^{\langle\sigma\rangle}. \tag{3}$$

Here E is predetermined and the locations of the errors are unknown. As in [17] we set *all* components of a vector $= \infty$ if $g(\xi_{1,\ell}, \ldots, \xi_{n,\ell}) = 0$, that even for those components with $f^{\langle\sigma\rangle}(\xi_{1,\ell}, \ldots, \xi_{n,\ell}) = 0$, but false vectors full of ∞'s can appear for $g(\xi_{1,\lambda_\kappa}, \ldots, \xi_{n,\lambda_\kappa}) \neq 0$. We can identify vectors that contain both ∞ and a field element as erroneous. Errors are dealt with by interpolating $(f^{\langle\sigma\rangle}\Lambda)/(g\Lambda)$ à la [15] where $\Lambda = (x_{n_1} - \xi_{n_1,\lambda_1}) \cdots (x_{n_1} - \xi_{n_1,\lambda_k})$ is an error locator polynomial for a chosen n_1 with $1 \leq n_1 \leq n$. We have the maximal supports

$$\left.\begin{array}{l} D_{f,E;n_1}^{\langle\sigma\rangle} = \{\tau x_{n_1}^\nu \mid \tau \in D_f^{\langle\sigma\rangle}, 0\leq\nu\leq E\} \supseteq \mathrm{supp}(f^{\langle\sigma\rangle}\Lambda), \\ D_{g,E;n_1} = \{\tau x_{n_1}^\nu \mid \tau \in D_g, 0\leq\nu\leq E\} \quad \supseteq \mathrm{supp}(g\Lambda). \end{array}\right\} \tag{4}$$

Now we limit the sparse supports of polynomials with unknown coefficients $\Phi^{\langle\sigma\rangle}$ and Ψ to the term sets (4). From (2) and (3) we obtain linear homogeneous equations for the coefficients of $\Phi^{\langle\sigma\rangle}$, Ψ:

$$\left.\begin{array}{l} \Phi^{\langle\sigma\rangle}(\xi_{1,\ell}, \ldots, \xi_{n,\ell}) - \beta_\ell^{\langle\sigma\rangle}\Psi(\xi_{1,\ell}, \ldots, \xi_{n,\ell}) = 0, \\ \quad \text{for } 0\leq\ell\leq L-1, 1\leq\sigma\leq s \text{ with } \beta_\ell^{\langle\sigma\rangle} \neq \infty, \\ \Psi(\xi_{1,\ell}, \ldots, \xi_{n,\ell}) = 0, \\ \quad \text{for } 0\leq\ell\leq L-1 \text{ with } \beta_\ell^{\langle 1\rangle} = \cdots = \beta_\ell^{\langle s\rangle} = \infty, \quad \text{with} \\ \mathrm{supp}(\Phi^{\langle\sigma\rangle}) \subseteq D_{f,E;n_1}^{\langle\sigma\rangle} \text{ for } 1\leq\sigma\leq s, \mathrm{supp}(\Psi) \subseteq D_{g,E;n_1}. \end{array}\right\} \tag{5}$$

Note that $\Phi^{\langle\sigma\rangle} \leftarrow f^{\langle\sigma\rangle}\Lambda$, $\Psi \leftarrow g\Lambda$ solve (5). We call any solution $(\Phi^{\langle 1\rangle}, \ldots, \Phi^{\langle s\rangle}, \Psi)$ of (5) an interpolant. We seek a (minimal) L and $\xi_{\mu,\ell}$ such that *all* solutions of (5) satisfy

$$\forall\sigma, 1 \le \sigma \le s\colon \Phi^{\langle\sigma\rangle} g = f^{\langle\sigma\rangle}\Psi,$$
$$\text{with } \operatorname{supp}(\Phi^{\langle\sigma\rangle}) \subseteq D^{\langle\sigma\rangle}_{f,E;n_1},\ \operatorname{supp}(\Psi) \subseteq D_{g,E;n_1}. \quad (6)$$

We call (6) the Welch-Berlekamp property. Then any non-zero solution vector to (6) satisfies

$$[\Phi^{\langle 1\rangle}/\Psi, ..., \Phi^{\langle s\rangle}/\Psi] = [f^{\langle 1\rangle}/g, ..., f^{\langle s\rangle}/g].$$

Theorem 2.1. *Let* $L{=}|D_{g,E;n_1}|{+}(\max_{1\le\sigma\le s}|D^{\langle\sigma\rangle}_{f,E;n_1}|){-}1$, $M^{\langle\sigma\rangle}{=}|D_{g,E;n_1}|{+}|D^{\langle\sigma\rangle}_{f,E;n_1}|$, *and let all* $\xi_{\mu,\ell}$, *where* $1 \le \mu \le n$ *and* $0 \le \ell \le L-1$, *be randomly and uniformly selected from a finite subset* $S \subseteq \mathsf{K}$. *Then the probability that all interpolant* $(s+1)$*-tuples* $(\Phi^{\langle 1\rangle}, ..., \Phi^{\langle s\rangle}, \Psi)$ *to* (5) *satisfy the Welch-Berlekamp property* (6) *is bounded from below as* $\ge 1 - (\sum_{\sigma=1}^{s}(M^{\langle\sigma\rangle} - E - 1)(\max\{\deg(\tau_f) \mid \tau_f \in D^{\langle\sigma\rangle}_f\} + \max\{\deg(\tau_g) \mid \tau_g \in D_g\} + E))/|S|$.

Because of page restrictions, we cannot detail the proof. We base the argument on 3 Lemmas. Our first auxiliary lemma relates solutions with the Welch-Berlekamp property (6) to interpolants of (5)

Lemma 2.2. *For any* $L \ge k \ge 0$, *any* $E \ge 0$ *and any evaluations* $\xi_{\mu,\ell} \in \mathsf{K}$, *where* $1 \le \mu \le n$, $0 \le \ell \le L-1$, *consider the solution* $(s+1)$*-tuples* $(\Phi^{\langle 1\rangle}, ..., \Phi^{\langle s\rangle}, \Psi)$ *to the homogeneous linear equations in their coefficients*

$$\forall\sigma, 1{\le}\sigma{\le}s\colon \Phi^{\langle\sigma\rangle} g{=}f^{\langle\sigma\rangle}\Psi \quad (\textit{that is, } (6)) \quad (7)$$
$$\Phi^{\langle\sigma\rangle}(\xi_{1,\lambda_\kappa}, ..., \xi_{n,\lambda_\kappa}){-}\beta^{\langle\sigma\rangle}_{\lambda_\kappa}\Psi(\xi_{1,\lambda_\kappa}, ..., \xi_{n,\lambda_\kappa}){=}0,$$
$$\textit{for } 1{\le}\kappa{\le}k, 1{\le}\sigma{\le}s \textit{ with } \beta^{\langle\sigma\rangle}_{\lambda_\kappa}{\ne}\infty, \quad (8)$$
$$\Psi(\xi_{1,\lambda_\kappa}, ..., \xi_{n,\lambda_\kappa}){=}0, \textit{for } 1{\le}\kappa{\le}k \textit{ with } \beta^{\langle 1\rangle}_{\lambda_\kappa}{=}\cdots{=}\beta^{\langle s\rangle}_{\lambda_\kappa}{=}\infty, \quad (9)$$
$$\Psi(\xi_{1,\ell}, ..., \xi_{n,\ell}){=}0,\ \textit{with } \ell{\notin}\{\lambda_1, ..., \lambda_k\} \textit{ and}$$
$$g(\xi_{1,\ell}, ..., \xi_{n,\ell}){=}0, \forall\sigma\colon f^{\langle\sigma\rangle}(\xi_{1,\ell}, ..., \xi_{n,\ell}){=}0. \quad (10)$$
$$\textit{with } \operatorname{supp}(\Phi^{\langle\sigma\rangle}){\subseteq}D^{\langle\sigma\rangle}_{f,E;n_1} \textit{ for } 1{\le}\sigma{\le}s, \operatorname{supp}(\Psi){\subseteq}D_{g,E;n_1}.$$

All those solution tuples must be interpolants of (5).

Our second auxiliary lemma gives an upper bound on L so that all interpolants of (5) for certain $\xi_{\mu,\ell}$ are in the described subspace of Lemma 2.2, meaning that they satisfy the Welch-Berlekamp property (6). The argument for $s = 1$ is already in [18, Section 4.1].

Lemma 2.3. *Let* $L_\times = |D_{g,E;n_1}| \times (\max_{1\le\sigma\le s}|D^{\langle\sigma\rangle}_{f,E;n_1}|)$ *and let* $\xi_{\mu,\ell} = \xi^\ell_\mu \in \mathsf{K}$, *where* $1 \le \mu \le n$ *and* $0 \le \ell \le L_\times - 1$, *such that for* $D^{\langle\sigma\rangle}_{f,E;n_1} \times D_{g,E;n_1} = \{\tau_f\tau_g \mid \tau_f \in D^{\langle\sigma\rangle}_{f,E;n_1}, \tau_g \in D_{g,E;n_1}\}$ *we have*

$$\tau_1(\xi_1, ..., \xi_n) \ne \tau_2(\xi_1, ..., \xi_n) \textit{ for all } \sigma \textit{ and}$$
$$\textit{for all } \tau_1, \tau_2 \in D^{\langle\sigma\rangle}_{f,E;n_1} \times D_{g,E;n_1}, \tau_1 \ne \tau_2, \quad (11)$$

(see [17, Assumption 4]). Then all interpolants $(s+1)$*-tuples* $(\Phi^{\langle 1\rangle}, ..., \Phi^{\langle s\rangle}, \Psi)$ *of* (5) *satisfy the Welch-Berlekamp property* (6).

Remark 2.1. For $n = 1$ and dense support sets $D^{\langle\sigma\rangle}_{f,E;n_1} = \{1, x_1, x_1^2, ...\}$, $D_{g,E;n_1} = \{1, x_1, x_1^2, ...\}$ we may choose $\xi_{1,\ell} = \hat{\xi}_\ell \in \mathsf{K}$ with $\hat{\xi}_{\ell_1} \ne \hat{\xi}_{\ell_2}$ for all $\ell_1 \ne \ell_2$. Then the coefficient matrix for $(\Phi^{\langle\sigma\rangle} g\Lambda - f^{\langle\sigma\rangle}\Lambda\Psi)(\hat{\xi}_\ell)$ is a non-zero Vandermonde matrix and the Lemma holds. □

The third lemma is the crucial idea that reduces $L_\times$ of Lemma 2.3. We will evaluate the black box for (1) at symbols $v_{\mu,\ell} \in \mathsf{K}(..., v_{\mu,\ell}, ...)$. It is not required from the black box to allow such elements (in transcendental extensions of K) as arguments, we solely use it for purpose of proof.

Lemma 2.4. *Let* $L_+ = |D_{g,E;n_1}| + (\max_{1\le\sigma\le s}|D^{\langle\sigma\rangle}_{f,E;n_1}|) - E - 1$. *Suppose* $\xi_{\mu,\ell} = v_{\mu,\ell}$ *is a new symbol (variable), for each* $1 \le \mu \le n$, $0 \le \ell \le L_+ - 1$. *We assume that* $k = 0$, *that is, there are no erroneous evaluations, so that* $\beta^{\langle\sigma\rangle}_\ell = (f^{\langle\sigma\rangle}/g)(v_{1,\ell}, ..., v_{n,\ell}) \in \mathsf{K}(v_{1,0}, ..., v_{n,L_+-1})$ *for all* ℓ. *Note that at vectors of* n *distinct variables there cannot be true poles. Then all interpolants* $(s+1)$*-tuples of* (5) *for* $L = L_+$ *over* $\mathsf{K}(v_{1,0}, ..., v_{n,L_+-1})$ *satisfy the Welch-Berlekamp property* (6).

The crucial idea for the proof of Lemma 2.4 is that by Lemma 2.3 for evaluations at $\xi_{\mu,\ell} = v^\ell_\mu$, where v_μ are fresh indeterminates (variables), the Welch-Berlekamp property (6) is attained from $L_\times$ evaluations. The property remains for evaluations $\xi_{\mu,\ell} = v_{\mu,\ell}$ because the rank of the corresponding linear system cannot drop: we can substitute $v_{\mu,\ell} = v^\ell_\mu$ in the coefficients to have at least the rank before. Finally, we can select a maximal set of linearly independent equations in (5) (for $\xi_{\mu,\ell} = v_{\mu,\ell}$). The equations at $\ell = \ell_\theta$ where $\theta = 1, 2, ...$ have a fresh set of variables so we can also use the initial $\ell = 1, 2, ...$. We can treat each σ individually; combining the equations (5) restricts the solutions to those with a common denominator Ψ.

The proof of Theorem 2.1 first uses for each σ separately the equations in (5) without erroneous $\beta^{\langle\sigma\rangle}_\ell$. We consider those $\xi_{\mu,\ell}$ which when substituted for $v_{\mu,\ell}$ in Lemma 2.4 preserve the rank, i.e., the dimensionality of the solution space. The complication with (10) is avoided by excluding such $\xi_{\mu,\ell}$. The solutions thus remain in the subspace of Lemma 2.2 (for $k = 0$). Erroneous equations and a common denominator restrict the solution space further, but there always is a solution $(f^{\langle 1\rangle}\Lambda, ..., f^{\langle s\rangle}\Lambda, g\Lambda)$. The probability of success is estimated by the Zippel-Schwartz Lemma. Note that at error locations our solutions satisfy

$$\forall\sigma\colon \Psi(\xi_{1,\lambda_\kappa}, ...) = 0 = \Phi^{\langle\sigma\rangle}(\xi_{1,\lambda_\kappa}, ...). \quad (12)$$

Remark 2.2. The condition $g \ne 0$ in (1) is not essential, which is useful when we inadvertently have projected the denominator to 0 in Section 3. If $g = 0$, all non-faulty evaluation vectors are by definition $[\infty, ..., \infty]$. If for all σ: $D^{\langle\sigma\rangle}_f = \varnothing \Rightarrow f^{\langle\sigma\rangle} = 0$, then the Welch-Berlekamp property (6) is satisfied for any solution (Φ, Ψ). Otherwise, if for one σ_1 we have $D^{\langle\sigma_1\rangle}_f \ne \varnothing$, we must have $\Psi = 0$ for all solutions of (5). The justification follows the original arguments, which are only partially presented above. □

3. VARIABLE-BY-VARIABLE SPARSE INTERPOLATION

In [7, 18] the variable-by-variable sparse reconstruction of [24] is generalized to multivariate rational functions, and in [17] an error-tolerant algorithm is introduced. We now modify the reconstruction in [17] so that an arbitrary error rate $< 1/2$ can be handled. Let

$$f/g \in \mathsf{K}(x_1, ..., x_n), \quad g \ne 0$$

be a rational function with a sparse numerator and denominator. We do not assume that $\mathrm{GCD}(f, g) = 1$, and wish to recover the sparse representations of f and g,

$$f{=}\sum_{j=1}^{t_f} a_j \vec{x}^{\vec{d}_j}, g{=}\sum_{m=1}^{t_g} b_m \vec{x}^{\vec{e}_m}, a_j, b_m{\in}\mathsf{K}, a_j{\ne}0, b_m{\ne}0, \quad (13)$$

the terms of the non-zero monomials are denoted by $\vec{\boldsymbol{x}}^{\vec{d}_j} = x_1^{d_{j,1}} \cdots x_n^{d_{j,n}}$ and $\vec{\boldsymbol{x}}^{\vec{e}_m} = x_1^{e_{m,1}} \cdots x_n^{e_{m,n}}$. We assume that a black box evaluates the function at $(x_1, ..., x_n) \leftarrow (\xi_1, ..., \xi_n) \in \mathsf{K}^n$, that with an adjusted error rate $1/q < 1/2$ (see Remark 1.1).

Lemma 3.1. *Suppose for all $\delta \geq 1$ and all $E \geq 0$, δ values are decoded from $\geq \delta + 2E$ evaluations at distinct points with $k \leq E$ errors. Then given δ and an adjusted error rate $1/q < 1/2$, one decodes correctly at any sequence of length $L \geq \max\{L_{\min}, \delta + 2\lfloor \delta/(q-2)\rfloor\}$ of evaluations at distinct points.*

Proof of Lemma 3.1. We have for $E = \lfloor \delta/(q-2)\rfloor$ that $E \leq \delta/(q-2) \Rightarrow E \leq (\delta+2E)/q$ and $(\delta+2E)/q \leq (\delta + 2\delta/(q-2))/q = \delta/(q-2)$, both of which imply that $E = \lfloor(\delta + 2E)/q\rfloor$. So a sequence of $\delta + 2E \geq L_{\min}$ evaluations with $k \leq (\delta+2E)/q$ errors is decoded. Now let $L \geq \delta + 2E + 1$. Then $L - 2\lfloor L/q\rfloor \geq (q-2)L/q \geq (q-2)(\delta+2E+1)/q = (q-2)(\delta+2\lfloor\delta/(q-2)\rfloor+1)/q \geq (q-2)(\delta+2(\delta/(q-2)-1)+1)/q = \delta - (q-2)/q > \delta - 1$. Since $L - 2\lfloor L/q\rfloor$ and δ are integers, $\delta \leq L - 2\lfloor L/q\rfloor$ values are decoded in the presence of $\leq \lfloor L/q\rfloor$ errors for all such $L \geq L_{\min}$. □

We now analyze a single iterative step. For that we assume that we have term sets

$$\left.\begin{aligned} D_{f,n-1} &\supseteq \{x_1^{d_{j,1}} x_2^{d_{j,2}} \cdots x_{n-1}^{d_{j,n-1}} \mid 1 \leq j \leq t_f\}, \\ D_{g,n-1} &\supseteq \{x_1^{e_{m,1}} x_2^{e_{m,2}} \cdots x_{n-1}^{e_{m,n-1}} \mid 1 \leq m \leq t_g\}. \end{aligned}\right\} \quad (14)$$

Those term sets come from having interpolated $(f/g)(x_1, ..., x_{n-1}, \alpha_n)$ for a random anchor point $\alpha_n \in S \subseteq \mathsf{K}$. In our algorithm we will iterate through the variables $x_2, ..., x_n$, starting with the dense $(f/g)(x_1, \alpha_2, ..., \alpha_n)$. We obtain the next term sets $D_{f,n}, D_{g,n}$ in 3 steps.

Step 1: We assume that on input we have upper bounds $\bar{d}_f^{[n]} \geq \deg_{x_n}(f) = \max_j\{d_{j,n}\}$ and $\bar{d}_g^{[n]} \geq \deg_{x_n}(g) = \max_m\{e_{m,n}\}$. For all $0 \leq \ell \leq L_{n-1} - 1$, where $L_{n-1} = |D_{f,n-1}| + |D_{g,n-1}| - 1$, (see (22) below), we select a random $(n-1)$-tuple $(\xi_{1,\ell}, ..., \xi_{n-1,\ell}) \in S^{n-1} \subseteq \mathsf{K}^{n-1}$ and interpolate the possibly unreduced univariate fraction $(f/g)(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n) \in \mathsf{K}(x_n)$. We use a univariate dense Cauchy interpolation algorithm with errors [15, 17]; see also Remark 2.1. We deal with the case $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n) = 0$ in the next paragraph. If we have $k_\ell \leq E$ errors in the batch of evaluations for each ℓ, then we use (correct and erroneous) values $\beta_{\ell,\ell'}$ of $(f/g)(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\ell'})$ for $0 \leq \ell' \leq L_\ell^{[n]} - 1$, where $L_\ell^{[n]} \leftarrow \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 2E + 1$ and where $\hat{\xi}_{\ell'_1} \neq \hat{\xi}_{\ell'_2}$ for all $\ell'_1 \neq \ell'_2$. Note that the list of evaluations points $\hat{\xi}_{\ell'}$ must be the same for all $(\xi_{1,\ell}, ..., \xi_{n-1,\ell})$. In the presence of an adjusted error rate $1/q < 1/2$, by Lemma 3.1 with $\delta = \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$ we can choose $L_\ell^{[n]} \leftarrow \max\{\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 2E + 1, L_{\min}\}$ with $E = \lfloor(\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1)/(q-2)\rfloor$. All computed solution pairs $(\Phi_\ell^{[n]}(x_n), \Psi_\ell^{[n]}(x_n)) \in \mathsf{K}[x_n]^2$ satisfy, by Theorem 2.1 as is explained in Remark 2.1, the Welch-Berlekamp property

$$f_\ell^{[n]}(x_n)\Psi_\ell^{[n]}(x_n) = \Phi_\ell^{[n]}(x_n) g_\ell^{[n]}(x_n), \quad (15)$$

where $f_\ell^{[n]}(x_n) = f(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n)$, $g_\ell^{[n]}(x_n) = g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n)$, and at all error locations λ' we have by (12) $\Phi_\ell^{[n]}(\hat{\xi}_{\lambda'}) = \Psi_\ell^{[n]}(\hat{\xi}_{\lambda'}) = 0$. A complication is that non-error locations can also have $\Phi_\ell^{[n]}(\hat{\xi}_{\ell'}) = \Psi_\ell^{[n]}(\hat{\xi}_{\ell'}) = 0$. Let $h = \mathrm{GCD}(f_\ell^{[n]}, g_\ell^{[n]})$; note that $g_\ell^{[n]}$ is assumed to be non-zero, for now. Then by (15) we have $\Phi_\ell^{[n]} = \hat{\Lambda} f_\ell^{[n]}/h$ and $\Psi_\ell^{[n]} = \hat{\Lambda} g_\ell^{[n]}/h$ for some polynomial $\hat{\Lambda} \in \mathsf{K}[x_n]$. Because none of the evaluation points with erroneous evaluations $\hat{\xi}_{\lambda'}$ can be roots of both of the relatively prime $f_\ell^{[n]}/h$ and $g_\ell^{[n]}/h$, the error locator polynomial $\Lambda = \prod_{\lambda'}(x_n - \hat{\xi}_{\lambda'})$ must divide $\hat{\Lambda}$. For a true pole $\hat{\xi}_{\ell'}$ ($\beta_{\ell,\ell'} = \gamma_{\ell,\ell'} = \infty \Rightarrow g_\ell^{[n]}(\hat{\xi}_{\ell'}) = 0$) with $(g_\ell^{[n]}/h)(\hat{\xi}_{\ell'}) \neq 0$ ($\Rightarrow h(\hat{\xi}_{\ell'}) = 0 \Rightarrow f_\ell^{[n]}(\hat{\xi}_{\ell'}) = 0$) the condition $\Psi_\ell^{[n]}(\hat{\xi}_{\ell'}) = 0$ implies that $x_n - \hat{\xi}_{\ell'}$ also divides $\hat{\Lambda}$. For each ℓ, we can compute a *lowest degree* non-zero interpolant pair $(\hat{\Phi}_\ell^{[n]}(x_n), \hat{\Psi}_\ell^{[n]}(x_n))$ and deduce an index set

$$\begin{aligned} I_\ell &= \{\ell' \mid \hat{\Phi}_\ell^{[n]}(\hat{\xi}_{\ell'}) = \hat{\Psi}_\ell^{[n]}(\hat{\xi}_{\ell'}) = 0\} \\ &= \{\ell' \mid \beta_{\ell,\ell'} \neq \gamma_{\ell,\ell'}\} \cup \{\ell' \mid f(\xi_{1,\ell}, ..., \\ &\qquad \xi_{n-1,\ell}, \hat{\xi}_{\ell'}) = g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\ell'}) = 0\}. \end{aligned} \quad (16)$$

Note that the sets in (16) above are not necessarily disjoint. We can also identify the corresponding correct evaluations

$$\beta_{\ell,\ell'} = \gamma_{\ell,\ell'} = (f/g)(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\ell'}), \qquad \text{for } \ell' = 0, 1, ..., L_\ell^{[n]} - 1, \ell' \notin I_\ell. \quad (17)$$

Because we have not assumed $\mathrm{GCD}(f, g) = 1$, we do not have the degrees $\deg_{x_n}(f)$ and $\deg_{x_n}(g)$ or the sparse supports in x_n of f and g. Note that we allow evaluations at poles $\gamma_{\ell,\ell'} = \infty$, but that it may be impossible for erroneous $\beta_{\ell,\lambda'} \neq \infty$ to identify a pole $\gamma_{\ell,\lambda'} = \infty$ when $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\lambda'}) = f(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\lambda'}) = 0$ ($\Rightarrow (f/g)(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n)$ is unreduced).

If $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n) = 0$, at least $L_\ell^{[n]} - \lfloor L_\ell^{[n]}/q\rfloor \geq \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + \lfloor L_\ell^{[n]}/q\rfloor + 1$ evaluations for $x_n \leftarrow \hat{\xi}_{\ell'}$ must yield ∞ (see Lemma 3.1 and also Remark 2.2). The observer, assuming $\leq \lfloor L_\ell^{[n]}/q\rfloor$ of those are false, still has $\geq \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1 \geq \bar{d}_g^{[n]} + 1$ true ∞'s, which by $\deg_{x_n}(g) \leq \bar{d}_g^{[n]}$ implies $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n) = 0$. We can deduce the correct evaluations $\gamma_{\ell,\ell'} = \infty$ for all ℓ', meaning $I_\ell = \varnothing$ in this case.

Step 2: Here we process each ℓ separately. For latter analysis, we need to have at least $\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$ known values $\gamma_{\ell,\ell'}$. In other words, we need to have a range $0 \leq \ell' \leq L_\ell^{[n]} - 1$ with $L_\ell^{[n]} - |I_\ell| \geq \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$. If $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n) = 0$ we have $I_\ell = \varnothing$, so our current range $L_\ell^{[n]}$ is sufficient. So let $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n) \neq 0$. The (unidentifiable) true poles $\hat{\xi}_{\ell'}$ with $\ell' \in I_\ell$ in (16) must be roots of $g(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, x_n)$, whose degree $\leq \bar{d}_g^{[n]}$. Hence $|I_\ell| \leq k_\ell + \bar{d}_g^{[n]} \leq \lfloor L_\ell^{[n]}/q\rfloor + \bar{d}_g^{[n]}$, thus if $\lfloor L_\ell^{[n]}/q\rfloor \geq \bar{d}_g^{[n]} \Rightarrow L_\ell^{[n]} - |I_\ell| \geq L_\ell^{[n]} - 2\lfloor L_\ell^{[n]}/q\rfloor \geq \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$, the latter by Lemma 3.1, and we are done. Note that $L_\ell^{[n]}$ has been determined from the adjusted error rate, incl. $L_{\min}$, and the degree bounds, all of which are known on input.

But it is possible that $L_\ell^{[n]} - |I_\ell| < \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$, that even for all ℓ, for example, when $f(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = g(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = 0$ for $\bar{d}_g^{[n]}$ of the $\hat{\xi}_{\ell'}$ and $L_\ell^{[n]}/q$ is too small. Separately for each such ℓ, we then increase the range of ℓ', computing additional $\beta_{\ell,\ell'}$ for reset $L_\ell^{[n]}$ and with them new I_ℓ. The new distinct evaluation points $\hat{\xi}_{\ell'}$ need not be random. For an adjusted error rate $1/q$, the process must stop at $L_\ell^{[n]} = \hat{L}_\ell^{[n]} = \hat{\delta} + \hat{E}$, where $\hat{\delta} = \bar{d}_f^{[n]} + 2\bar{d}_g^{[n]} + 1$ and $\hat{E} = \lfloor\hat{\delta}/(q-1)\rfloor$

$= \lfloor \hat{L}_\ell^{[n]}/q \rfloor$: we had $E{=}\lfloor(\bar{d}_f^{[n]}{+}\bar{d}_g^{[n]}{+}1)/(q{-}2)\rfloor$ (see Step 1) $\le \lfloor L_\ell^{[n]}/q\rfloor$ (because $E{\le}(\bar{d}_f^{[n]}{+}\bar{d}_g^{[n]}{+}2E{+}1)/q{\le}L_\ell^{[n]}/q$; see Lemma 3.1) $< \bar{d}_g^{[n]}$ (otherwise, we were done), which implies $\bar{d}_f^{[n]}{+}\bar{d}_g^{[n]}{+}1{<}(q{-}2)\bar{d}_g^{[n]} \Rightarrow \hat{E}{<}\bar{d}_g^{[n]}$. So $\le\lfloor \hat{L}_\ell^{[n]}/q\rfloor{\le}\hat{E}$ errors are indeed identified from $\hat{L}_\ell^{[n]}{>}\bar{d}_f^{[n]}{+}\bar{d}_g^{[n]}{+}2\hat{E}{+}1$ evaluations, and for the corresponding $\hat{I}_\ell$ we have $\hat{L}_\ell^{[n]}{-}|\hat{I}_\ell| \ge \hat{L}_\ell^{[n]}{-}(\bar{d}_g^{[n]}{+}\hat{E}) = \bar{d}_f^{[n]}{+}\bar{d}_g^{[n]}{+}1$.

One also may prevent, with high probability, the necessity of enlarging the range of ℓ' by randomly selecting the first $\max\{\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + E + 1, L_{\min}\}$ of the $\hat{\xi}_{\ell'} \in S'$. Probabilistic failure at this stage can be diagnosed and corrected by enlarging the range as above with new random choices $\hat{\xi}_{\ell'} \in S'$.

Step 3: We now use the good evaluations (17) and the additional values from Step 2 to obtain the sparse solutions (Φ, Ψ). We solve, for the coefficients of Φ, Ψ, the following linear homogeneous system:

$$\left.\begin{array}{l}
\Phi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}){-}\gamma_{\ell,\ell'}\Psi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}){=}0,\\
\text{for } 0{\le}\ell{\le}L_{n-1}{-}1, 0{\le}\ell'{\le}L_\ell^{[n]}{-}1, \ell'{\notin}I_\ell, \text{with } \gamma_{\ell,\ell'}{\ne}\infty,\\
\Psi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}){=}0,\\
\text{for } 0{\le}\ell{\le}L_{n-1}{-}1, 0{\le}\ell'{\le}L_\ell^{[n]}{-}1, \ell'{\notin}I_\ell, \text{with } \gamma_{\ell,\ell'}{=}\infty,\\
\text{with } \operatorname{supp}(\Phi) \subseteq \bar{D}_{f,n}{=}\{\tau_f x_n^\delta | \tau_f{\in}D_{f,n-1}, 0{\le}\delta{\le}\bar{d}_f^{[n]}\},\\
\qquad \operatorname{supp}(\Psi) \subseteq \bar{D}_{g,n}{=}\{\tau_g x_n^\eta | \tau_g{\in}D_{f,n-1}, 0{\le}\eta{\le}\bar{d}_g^{[n]}\}.
\end{array}\right\} \quad (18)$$

Note that L_{n-1} is chosen (see (22) below) so that the Cauchy property $f\,\Psi = \Phi\,g$, namely, the Welch-Berlekamp property without errors, is satisfied for all solutions (Φ, Ψ), that with high probability. The difficulty in applying Theorem 2.1 (for $s = 1$, $E = 0$ and $n_1 = n$) is that the random $(\xi_{1,\ell}, \dots, \xi_{n-1,\ell})$ are repeated for the $\hat{\xi}_{\ell'}$ over the range of ℓ', and therefore the arguments $(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}, \hat{\xi}_{\ell'})$ would not be completely random for each equation in (18), even if the $\hat{\xi}_{\ell'}$ were chosen randomly.

First, we argue that any solution pair (Φ, Ψ) of (18) will satisfy the above equations for all $\ell' \in I_\ell$ and all ℓ as well, almost, namely,

$$\begin{array}{l}
\Phi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}){-}\gamma_{\ell,\ell'}\Psi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}) = 0,\\
\quad \text{for } 0{\le}\ell{\le}L_{n-1}{-}1, \ell' \in I_\ell, \text{ with } \gamma_{\ell,\ell'} \ne \infty, \qquad (19)\\
\Psi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}) = 0, \text{ for } 0{\le}\ell{\le}L_{n-1}{-}1,\\
\quad \ell'{\in}I_\ell \text{ with } \gamma_{\ell,\ell'}{=}\infty, f(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}){\ne}0, \qquad (20)\\
\quad \text{with } \operatorname{supp}(\Phi) \subseteq \bar{D}_{f,n}, \operatorname{supp}(\Psi) \subseteq \bar{D}_{g,n}.
\end{array}$$

Here (20) excludes the case $f(\xi_{1,\ell},\dots,\xi_{n-1,\ell},\hat{\xi}_{\ell'}) = g(\xi_{1,\ell}, \dots, \xi_{n-1,\ell},\hat{\xi}_{\ell'}) = 0$. As said earlier, equations for such poles cannot be deduced. The claim (18) $\Rightarrow$ (19,20) for each ℓ is non-trivial only if $I_\ell \ne \varnothing$, so we can assume that we have $g(\xi_{1,\ell},\dots,\xi_{n-1,\ell},x_n) \ne 0$. Then for each ℓ, we shall argue that the equations (18) produce (Φ, Ψ) such that their univariate projections $\Phi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},x_n), \Psi(\xi_{1,\ell},\dots,\xi_{n-1,\ell},x_n)$ satisfy the univariate Cauchy property

$$(f\Psi - \Phi g)(\xi_{1,\ell},\dots,\xi_{n-1,\ell},x_n) = 0; \qquad (21)$$

see Remark 2.1 for $n = s = 1$, $E = 0$, $D_{f,0;1}^{\langle 1\rangle} = \{1, x_n, x_n^2, \dots, x_n^{\bar{d}}\}$, $\bar{d} = \bar{d}_f^{[n]}$, and $D_{g,0;1} = \{1, x_n, x_n^2, \dots, x_n^{\bar{e}}\}$, $\bar{e} = \bar{d}_g^{[n]}$. The univariate Cauchy property (21) requires, by Theorem 2.1 and Remark 2.1, at least $\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$ distinct points $\hat{\xi}_{\ell'}$ in our system (18), including (true) poles. In Step 2 each I_ℓ was individually adjusted to that, namely $L_\ell^{[n]} - |I_\ell| \ge \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$. Then plugging $x_n \leftarrow \hat{\xi}_{\ell'}$ in the univariate Cauchy property (21) yields (19,20).

We now wish to apply Theorem 2.1 with $E = 0$ and $s = 1$ to the $(n{-}1)$-variate interpolation problem $(f/g)(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'})$. We fix ℓ', and analyze the solutions (Φ, Ψ) of the corresponding subset of equations in (18,19,20) for random $(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}) \in S^{n-1}$ for all ℓ,

$$0 \le \ell \le L_{n-1} - 1, \quad L_{n-1} = |D_{f,n-1}| + |D_{g,n-1}| - 1. \quad (22)$$

If $g(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'}) \ne 0$, the above equations (18,19,20) then constitute the set of interpolation equations (5) for $E = 0$ and $s = 1$ with one exception: the constraint $\Psi(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}, \hat{\xi}'_\ell) = 0$ is missing in (20) when $f(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}, \hat{\xi}'_\ell) = g(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}, \hat{\xi}'_\ell) = 0$. As stated in the sketch of the proof of Theorem 2.1 our random points $\xi_{1,0}, \dots, \xi_{n-1,\hat{L}}$, $\hat{L} = L_{n-1} - 1$, avoid condition (10) at the non-erroneous evaluation points that are required to attain the Welch-Berlekamp property. If additional constraining equations are missing, the Welch-Berlekamp property cannot be lost. Note that $g(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}, \hat{\xi}'_\ell) = 0$ while for some $\Psi(\xi_{1,\ell}, \dots, \xi_{n-1,\ell}, \hat{\xi}'_\ell) \ne 0$ because g need not divide Ψ since f/g is not reduced.

If $g(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'}) = 0$, Remark 2.2 becomes applicable only if the equations for the poles are not missing. We will address this case below, and for the moment assume that $g(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'}) \ne 0$ for all ℓ', $0 \le \ell' \le \bar{d}_f^{[n]} + \bar{d}_g^{[n]}$. Now the conclusions of Theorem 2.1 apply to our solutions (Φ, Ψ), as was explained in the previous paragraph. We thus have, for each ℓ', the $(n-1)$-variate Cauchy property

$$(f\Psi)(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'}) = (\Phi g)(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'}), \qquad (23)$$

that with probability $1 - p_{\ell'}$, where

$$p_{\ell'} \le p = \Big(L_{n-1}(\max\{\deg(\tau_f) \mid \tau_f \in D_{f,n-1}\} + \max\{\deg(\tau_g) \mid \tau_g \in D_{g,n-1}\})\Big)/|S|. \quad (24)$$

The probability that for the solution pair (Φ, Ψ) (23) is invalid for one or more of the ℓ', $0 \le \ell' \le \bar{d}_f^{[n]} + \bar{d}_g^{[n]}$, is $\le \sum_{\ell'} p_{\ell'} \le (\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1)p$ (the events can be dependent), so (23) is valid for all such ℓ' with probability $\ge 1 - (\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1)p$.

Now suppose that a pair of non-zero polynomials $\Phi = \sum_{\tau_f \in D_{f,n-1}} \Xi_{\tau_f}(x_n)\tau_f$ and $\Psi = \sum_{\tau_g \in D_{g,n-1}} \Omega_{\tau_g}(x_n)\tau_g$, with $\Xi_{\tau_f}(x_n), \Omega_{\tau_g}(x_n) \in \mathsf{K}[x_n]$, satisfies the $(n{-}1)$-variate Cauchy property (23) for all ℓ', where $0 \le \ell' \le \bar{d}_f^{[n]} + \bar{d}_g^{[n]}$. Let $f = \sum_{\tau_f \in D_{f,n-1}} A_{\tau_f}(x_n)\tau_f$ and $g = \sum_{\tau_g \in D_{g,n-1}} B_{\tau_g}(x_n)\tau_g$, with $A_{\tau_f}(x_n), B_{\tau_g}(x_n) \in \mathsf{K}[x_n]$; note that by the superset relation in (14), some $A_{\tau_f}(x_n)$ and $B_{\tau_g}(x_n)$ can be zero. We now have, for all those ℓ',

$$0 = (f\Psi - \Phi g)(x_1, \dots, x_{n-1}, \hat{\xi}_{\ell'}) = \sum_\tau \Big(\sum_{\tau_f, \tau_g :\, \tau_f\tau_g = \tau} A_{\tau_f}(\hat{\xi}_{\ell'})\Omega_{\tau_g}(\hat{\xi}_{\ell'}) - \Phi_{\tau_f}(\hat{\xi}_{\ell'})B_{\tau_g}(\hat{\xi}_{\ell'})\Big)\tau,$$

hence all univariate polyn. coefficients $\sum_{\tau_f\tau_g=\tau} A_{\tau_f}(x_n) \times \Omega_{\tau_g}(x_n) - \Phi_{\tau_f}(x_n)B_{\tau_g}(x_n)$ vanish on at least $\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1 > \deg(A_{\tau_f}\Omega_{\tau_g} - \Phi_{\tau_f}B_{\tau_g})$ distinct points $\hat{\xi}_{\ell'}$ and therefore must be identically zero. We conclude that with probability $\ge 1 - (\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1)p$ our solutions to (18) must satisfy the Cauchy property $f\Psi - \Phi g = 0$.

We finally discuss the assumed condition $g(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) \neq 0$. In the previous paragraphs, we have needed $\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$ many such $\hat{\xi}_{\ell'}$. In the worst case, there are at most $\bar{d}_g^{[n]}$ values for $\hat{\xi}_{\ell'}$ that have $g(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = 0$ ($g \neq 0$), but we actually have $\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 2E + 1$ distinct $\hat{\xi}_{\ell'}$. If $E \geq \bar{d}_g^{[n]}/2$, we have sufficiently many $\hat{\xi}_{\ell'}$ with $g(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) \neq 0$ for our arguments in the paragraph before.

If not, we can simply allow $g(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = 0$. Then if also $f(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = 0$ we have the needed $(n-1)$-variate Cauchy property (23) for that ℓ'. Lastly, let $f(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) \neq 0$ ($\Rightarrow D_{f,n-1} \neq \varnothing$). Then the probability that for all ℓ, $0 \leq \ell \leq L_{n-1} - 1$, $f(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\ell'}) \neq 0$ is $\geq 1 - L_{n-1}(\max\{\deg(\tau_f) \mid \tau_f \in D_{f,n-1}\})/|S|$. Under this condition, with (18) $\Rightarrow$ (20) we have

$$\forall 0 \leq \ell \leq L_{n-1} - 1\colon \Psi(\xi_{1,\ell}, ..., \xi_{n-1,\ell}, \hat{\xi}_{\ell'}) = 0. \qquad (25)$$

Now Remark 2.2 applies: if the random $\xi_{1,0}, ..., \xi_{n-1,\hat{L}} \in S$, $\hat{L} = L_{n-1} - 1$, also preserve the maximal generic rank of the (projected) coefficient matrix of (25) right above, which is exactly $|D_{g,n-1}| \geq 1$, then $\Psi(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = 0$ and the $(n-1)$-variate Cauchy property (23) for that ℓ' is satisfied. The probability of $\xi_{1,0}, ..., \xi_{n-1,\hat{L}}$ being a root of a corresponding determinant is $\leq (|D_{g,n-1}| \max\{\deg(\tau_g) \mid \tau_g \in D_{g,n-1}\})/|S|$. By $|D_{f,n-1}| \geq 1$ we have $|D_{g,n-1}| \leq L_{n-1}$, and therefore one or both of the conditions fail with probability $\leq p$ (see (24)). Hence we obtain the $(n-1)$-variate Cauchy property (23) for an exceptional ℓ' with $g(x_1, ..., x_{n-1}, \hat{\xi}_{\ell'}) = 0$ with probability $\geq 1 - p$ as well.

In summary, with $\leq \max\{\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1 + \max\{\hat{E} + \bar{d}_g^{[n]}, 2E\}, L_{\min}\}L_{n-1}$ evaluations in the presence of an adjusted error rate $1/q < 1/2$, where $E = \lfloor(\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1)/(q-2)\rfloor$ and $\hat{E} = \lfloor(\bar{d}_f^{[n]} + 2\bar{d}_g^{[n]} + 1)/(q-1)\rfloor$, using $(n-1)L_{n-1}$ random field elements $\in S$, all computed solutions (Φ, Ψ) satisfy $f\Psi = \Phi g$ with probability

$$\geq 1 - \Big((\bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1)L_{n-1}(\max\{\deg(\tau_f) \mid \tau_f \in D_{f,n-1}\} + \max\{\deg(\tau_g) \mid \tau_g \in D_{g,n-1}\})\Big)/|S|.$$

4. NUMERICAL INTERPOLATION WITH A HIGH ERROR RATE

In this section, we provide the numerical approach based on the variable-by-variable sparse interpolation of Section 3 to recover sparse rational function from values with noise and outlier errors, where the *adjusted* error rate is $< 1/2$. Similar to [17], in the approximate case, a threshold Θ is introduced to separate an evaluation that is an outlier error, that is, if the evaluation β at the point $(\zeta_1, ..., \zeta_n) \in \mathbb{C}^n$ is an outlier error, then $\beta = \gamma + \gamma'$, where $\gamma = f(\zeta_1, ..., \zeta_n)/g(\zeta_1, ..., \zeta_n) \in \mathbb{C} \cup \{\infty\}$, and $|\gamma'/\gamma| \geq \Theta$; here false poles and non-poles are also allowed.

Consider the rational function $f/g \in \mathbb{C}(x_1, ..., x_n)$, where f, g are represented as (13). Suppose a black box for f/g with noise and outlier errors at a known *adjusted* error rate is given. Based on the univariate Cauchy interpolation algorithm [15, 17], we will propose a method to interpolate f and g variable by variable, i.e., recover f_i and g_i when f_{i-1} and g_{i-1} are obtained. More specifically, we will discuss how to generalize the method proposed in Section 3, including three steps, when the black box for f/g is with noise and outlier errors. Similarly, here we only consider a single iterative step, that is, how to interpolate $f(x_1, ..., x_n)$ and $g(x_1, ..., x_n)$ when term sets $D_{f,n-1}$ and $D_{g,n-1}$ of $f(x_1, ..., x_{n-1}, \alpha_n)$ and $g(x_1, ..., x_{n-1}, \alpha_n)$ are computed.

We at first generalize **Step 1** in Section 3 to the numerical case, i.e., apply the univariate numerical Cauchy algorithm [17] to identify the approximate evaluations without outlier errors from values with noise and outlier errors. Suppose $(f^{[n]}/g^{[n]})(x_n) = f/g(\alpha_1, ..., \alpha_{n-1}, x_n)$, and assume that the degree upper bounds of $\bar{d}_f \geq \deg_{x_n}(f)$ and $\bar{d}_g \geq \deg_{x_n}(g)$ are given. Note that it is unnecessary to require $f^{[n]}$ and $g^{[n]}$ are approximately relatively prime. According to Lemma 3.1, in order to recover $\Phi^{[n]}$ and $\Psi^{[n]}$ which satisfy the Welch-Berlekamp property, the number of evaluations can be determined by $L^{[n]} \leftarrow \max\{\bar{d}_f + \bar{d}_g + 2E + 1, L_{\min}\}$, where $E = \lfloor(\bar{d}_f + \bar{d}_g + 1)/(q-2)\rfloor$. Given a random root of unity $\zeta \in \mathbb{C}$, we get the evaluations with noise and outlier errors from the black box

$$\beta_{\ell'} = \gamma_{\ell'} + \gamma'_{\ell'}, \text{ where } \gamma_{\ell'} = \frac{f(\alpha_1, ..., \alpha_{n-1}, \zeta^{\ell'})}{g(\alpha_1, ..., \alpha_{n-1}, \zeta^{\ell'})} \in \mathbb{C} \cup \{\infty\}, \quad \ell' = 0, 1, ..., L^{[n]} - 1, \qquad (26)$$

where $\gamma'_{\ell'}$ denotes noise or possibly an outlier error. Moreover, the number of ℓ' such that $|\gamma'_{\ell'}/\gamma_{\ell'}| \geq \Theta$, is $\leq E$. From the evaluations (26), we show how to compute the interpolants $\Phi^{[n]}(x_n)$ and $\Psi^{[n]}(x_n)$ satisfy the Welch-Berlekamp property

$$f^{[n]}(x_n)\Psi^{[n]}(x_n) = \Phi^{[n]}(x_n)g^{[n]}(x_n).$$

Let $\vec{\boldsymbol{y}}$ and $\vec{\boldsymbol{z}}$ be the coefficient vectors of $\Phi^{[n]}(x_n)$ and $\Psi^{[n]}(x_n)$, respectively. From (26), we construct the following linear equations for $\ell' = 0, 1, ..., L^{[n]} - 1$,

$$\sum_{j=0}^{\bar{d}_f + E} y_j \zeta^{\ell' j} - \beta_{\ell'} \sum_{m=0}^{\bar{d}_g + E} z_m \zeta^{\ell' m} = 0. \qquad (27)$$

The above equations form a linear system

$$G \begin{bmatrix} \vec{\boldsymbol{y}} & \vec{\boldsymbol{z}} \end{bmatrix}^T = [V, -\Gamma W] \begin{bmatrix} \vec{\boldsymbol{y}} & \vec{\boldsymbol{z}} \end{bmatrix}^T = \mathbf{0}, \qquad (28)$$

where $\Gamma = \mathrm{diag}(\beta_0, \beta_1, ..., \beta_{L^{[n]}-1})$, and where V, W are Vandermonde matrices generated by the vectors $[1, \zeta, ..., \zeta^{\bar{d}_f + E}]^T$ and $[1, \zeta, ..., \zeta^{\bar{d}_g + E}]^T$. The numerical rank deficiency of G, denoted by ρ, can be computed by checking the number of small singular values of G or finding the largest gap among the singular values. Suppose $h(x_n)$ is the approximate GCD of $f^{[n]}(x_n)$ and $g^{[n]}(x_n)$ to the given tolerance, and denote $s = \min(\bar{d}_f - \deg_{x_n}(f), \bar{d}_g - \deg_{x_n}(g))$. According to the discussion in Section 3, we have $\rho = 1 + E - k + s + \deg(h)$. Once ρ is determined, the linear equations (27) are transformed into the following reduced linear equations by removing some columns corresponding to higher degree in (27), namely, for $\ell' = 0, 1, ..., L^{[n]} - 1$

$$\sum_{j=0}^{\bar{d}_f + E - \rho + 1} y_j \zeta^{\ell' j} - \beta_{\ell'} \sum_{m=0}^{\bar{d}_g + E - \rho + 1} z_m \zeta^{\ell' m} = 0, \qquad (29)$$

whose matrix form is

$$\widetilde{G} \begin{bmatrix} \vec{\boldsymbol{y}} & \vec{\boldsymbol{z}} \end{bmatrix}^T = [\widetilde{V}, -\Gamma \widetilde{W}] \begin{bmatrix} \vec{\boldsymbol{y}} & \vec{\boldsymbol{z}} \end{bmatrix}^T = \mathbf{0}. \qquad (30)$$

Clearly, the numerical rank deficiency of $\widetilde{G}$ is 1. The coefficient vector $\vec{\boldsymbol{y}}$ of $\Phi^{[n]}$ and the coefficient vector $\vec{\boldsymbol{z}}$ of $\Psi^{[n]}$ are achieved from the last singular vector of $\widetilde{G}$. For a preset tolerance $\epsilon_{\rm root}$, we can obtain an index set I (similar to (16) in the exact case), by checking for $\ell' = 0, 1, ..., L^{[n]} - 1$,

$$I = \{\ell' \mid |\Phi^{[n]}(\zeta^{\ell'})| + |\Psi^{[n]}(\zeta^{\ell'})| \le \epsilon_{\rm root}\} = \{\ell' \mid |\gamma'_{\ell'}/\gamma_{\ell'}| \ge \Theta\} \cup \{\ell' \mid |f^{[n]}(\zeta^{\ell'})| + |g^{[n]}(\zeta^{\ell'})| \le \epsilon_{\rm root}\}. \quad (31)$$

We can assume that I contains all error locations and the locations ℓ' for which $x_n - \zeta^{\ell'}$ is the approximate common divisor of $f^{[n]}$ and $g^{[n]}$. As discussed in the exact case, it is impossible to determine the exact error location set. However, we do obtain correct locations for approximate evaluations that are not outliers, and their values

$$\beta_{\ell'} \approx \gamma_{\ell'} = (f^{[n]}/g^{[n]})(\zeta^{\ell'}), \text{ for } \ell' = 0, 1, ..., L^{[n]} - 1, \ell' \notin I. \quad (32)$$

In other words, the above method is able to remove the outliers from the values with noise and outlier errors.

Remark 4.1. Described as **Step 2** in Section 3, once $L^{[n]} - |I| < \bar{d}_f^{[n]} + \bar{d}_g^{[n]} + 1$, we also need to increase the range of ℓ' and compute more approximate evaluations without outlier errors at the points $\zeta^{\ell'}$.

Having the above procedure to obtain the approximate evaluations without outlier errors, we now discuss how to compute the sparse solutions (Φ, Ψ) which satisfy the Welch-Berlekamp property, when the actual supports $D_{f,n-1}$ and $D_{g,n-1}$ are computed. Similar to (18), we construct the possible terms $\bar{D}_{f,n}$ and $\bar{D}_{g,n}$ of Φ and Ψ respectively from $D_{f,n-1}, D_{g,n-1}$. Suppose the possible terms $\bar{D}_{f,n}$ and $\bar{D}_{g,n}$ in Φ and Ψ are $\bar{D}_{f,n} = \{x_1^{\bar{d}_{j,1}} \cdots x_n^{\bar{d}_{j,n}} \mid j = 1, 2, ..., \bar{t}_f\}$ and $\bar{D}_{g,n} = \{x_1^{\bar{e}_{m,1}} \cdots x_n^{\bar{e}_{m,n}} \mid m = 1, 2, ..., \bar{t}_g\}$. The unknown polynomials Φ and Ψ are represented as $\Phi = \sum_{j=1}^{\bar{t}_f} y_j x_1^{\bar{d}_{j,1}} \cdots x_n^{\bar{d}_{j,n}}$, $\Psi = \sum_{m=1}^{\bar{t}_g} z_m x_1^{\bar{e}_{m,1}} \cdots x_n^{\bar{e}_{m,n}}$, where y_j and z_m are unknown.

For $0 \le \ell \le L_{n-1} - 1$ with $L_{n-1} = |D_{f,n-1}| + |D_{g,n-1}| - 1$, let $b_{1,\ell}, ..., b_{n-1,\ell} \in \mathbb{Z}_{>0}$ be sufficient large distinct prime numbers and $s_{j,\ell}$ be random integers with $1 \le s_{j,\ell} < b_{j,\ell}$, $1 \le j \le n-1$. For all ℓ, we choose a random $(n-1)$-tuple $(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell})$, where $\zeta_{j,\ell} = \exp(2\pi \boldsymbol{i}/b_{j,\ell})^{s_{j,\ell}} \in \mathbb{C}$, $1 \le j \le n-1$ (cf. [9]). For each ℓ, we choose a random root of unity $\hat{\zeta}_\ell$, and get the evaluations with noise and outlier errors

$$\beta_{\ell,\ell'} = \gamma_{\ell,\ell'} + \gamma'_{\ell,\ell'}, \text{ where } \gamma_{\ell,\ell'} = \frac{f(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell}, \hat{\zeta}_\ell^{\ell'})}{g(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell}, \hat{\zeta}_\ell^{\ell'})} \in \mathbb{C} \cup \{\infty\}, \ell' = 0, 1, ..., L_\ell^{[n]} - 1, \quad (33)$$

where $\gamma'_{\ell,\ell'}$ denotes noise or possibly an outlier error.

For each $\ell, 0 \le \ell \le L_{n-1} - 1$, from the evaluations $\beta_{\ell,\ell'}, 0 \le \ell' \le L_\ell^{[n]}$, one is able to apply the above procedure to get an index set I_ℓ including all error locations,

$$I_\ell = \{\ell' \mid |\beta_{\ell,\ell'}/\gamma_{\ell,\ell'}| \ge \Theta\} \cup \{\ell' \mid |f(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell}, \hat{\zeta}_\ell^{\ell'})| + |g(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell}, \hat{\zeta}_\ell^{\ell'})| \le \epsilon_{\rm root}\}. \quad (34)$$

By removing all the evaluations at I_ℓ for each ℓ, one obtains the approximate evaluations without outlier errors, that is,

$$\beta_{\ell,\ell'} \approx \frac{f(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell}, \hat{\zeta}_\ell^{\ell'})}{g(\zeta_{1,\ell}, ..., \zeta_{n-1,\ell}, \hat{\zeta}_\ell^{\ell'})}, \ell' = 0, 1 ..., L_\ell^{[n]} - 1, \ell' \notin I_\ell, \quad (35)$$

for $\ell = 0, ..., L_{n-1} - 1$. With the unknown y_j and z_m, according to (18) we construct the following linear system from the approximate evaluations (35),

$$G \begin{bmatrix} \vec{\boldsymbol{y}} & \vec{\boldsymbol{z}} \end{bmatrix}^T = [V, -\Gamma W] \begin{bmatrix} \vec{\boldsymbol{y}} & \vec{\boldsymbol{z}} \end{bmatrix}^T = \boldsymbol{0}, \quad (36)$$

where Γ is a diagonal matrix composed of $\beta_{\ell,\ell'}$ in (35). Presented in [18], a structured total least norm (STLN) method is applicable to compute the optimal deformation of Γ in (36). More details are found in [18].

Algorithm *Numerical Interpolation of Rational Functions with Outlier Errors*

Input: ▸ $\frac{f(x_1,...,x_n)}{g(x_1,...,x_n)} \in \mathbb{C}(x_1, \ldots, x_n)$ input as a black box with noise and outlier errors, an *adjusted* error rate.

▸ $(x_1, \ldots, x_n)$: an ordered list of variables in f/g.

▸ $\bar{d}_f, \bar{d}_g$: total degree bounds $\bar{d}_f \ge \deg(f)$ and $\bar{d}_g \ge \deg(g)$.

▸ $\epsilon_{\rm coeff} > 0$ (for "forcing underflow" of terms), $\epsilon_{\rm root} > 0$ (for zero detection), $\epsilon_{\rm rank} > 0$ (for numeric rank detection), the given tolerance.

Output: $f(x_1, \ldots, x_n)/c$ and $g(x_1, \ldots, x_n)/c$, where $c \in \mathbb{C}$.

1. Initialize the anchor points and the support of f and g: choose $\alpha_1, \alpha_2, \ldots, \alpha_n$ as random roots of unity, let $D_{f,0} = \{1\}$ and $D_{g,0} = \{1\}$.
2. For $i = 1, 2, \ldots, n$ do: Interpolate the polynomials f_i and g_i as follows:

(a) For $\ell = 0, 1, \ldots, |D_{f,i-1}| + |D_{g,i-1}| - 2$ do: Obtain the approximate evaluations without outlier errors.

(a.1) Choose random roots of unity $\zeta_{\ell,1}, \ldots, \zeta_{\ell,i-1}, \hat{\zeta}_\ell$.

(a.2) Get the evaluations $\beta_{\ell,\ell'}$ with the noise and the outlier errors as (33).

(a.3) Construct the matrix G in (28) from $\beta_{\ell,\ell'}$ and $\hat{\zeta}_\ell$. Compute the SVD of G and find its numerical rank deficiency ρ. Alternatively, a relative tolerance $\epsilon_{\rm rank}$ for a jump in the singular values could be provided as an additional input.

(a.4) Get the matrix $\widetilde{G}$ from the reduced linear system (29) with ρ, and then obtain $\Phi_\ell^{[i]}$ and $\Psi_\ell^{[i]}$ from the last singular vector of $\widetilde{G}$.

(a.5) Get an index set I_ℓ including the error locations by checking (34), and then obtain the approximate evaluations without outlier errors as (35).

(b) From $D_{f,i-1}$ and $D_{g,i-1}$, get the possible terms $\bar{D}_{f,i}$ and $\bar{D}_{g,i}$.

(c) From the approximate evaluations $\beta_{\ell,\ell'}$ without outlier errors from Step (a.5), and $\bar{D}_{f,i}, \bar{D}_{g,i}$, construct the linear system as in (18,36).

(d) Apply the STLN method to interpolate f_i and g_i, and then get their actual supports $D_{f,i}$ and $D_{g,i}$.

3. With the actual support $D_{f,n}$ and $D_{g,n}$ of f_n and g_n, interpolate $f(x_1, \ldots, x_n)/c$ and $g(x_1, \ldots, x_n)/c$ again to improve the accuracy of the coefficients:

(a) Construct the linear system from the approximate evaluations without outlier errors $\beta_{\ell,\ell'}$ as (18,36) and $D_{f,n}, D_{g,n}$.

(b) Compute the refined solution $\vec{\boldsymbol{y}}$ and $\vec{\boldsymbol{z}}$ by use of STLN method.

(c) Obtain $f(x_1, \ldots, x_n)/c$ and $g(x_1, \ldots, x_n)/c$ from $\vec{\boldsymbol{y}}, \vec{\boldsymbol{z}}$ and $D_{f,n}$, $D_{g,n}$. □

5. EXPERIMENTS

Our algorithm has been implemented in Maple and the performance is reported in the following two tables. All examples in Table 1 and Table 2 are run in Maple 15 under

E x	Random Noise	$\bar{d}_f, \bar{d}_g$	deg(f), deg(g)	t_f, t_g	n	1/q	N	Time (secs.)	Rel. Error
1	$10^{-5}\sim 10^{-3}$	3, 3	1, 1	1, 3	2	1/3	244	1.638	2.64e–7
2	$10^{-5}\sim 10^{-3}$	5, 5	2, 2	3, 3	2	1/3	336	2.621	2.80e–7
3	$10^{-5}\sim 10^{-3}$	2, 5	1, 4	2, 4	3	1/4	432	3.744	4.30e–7
4	$10^{-5}\sim 10^{-3}$	8, 8	5, 2	10, 6	3	1/4	507	13.31	2.87e–9
5	$10^{-6}\sim 10^{-4}$	10, 10	7, 7	10, 10	5	1/4	2193	127.7	8.14e–11
6	$10^{-7}\sim 10^{-5}$	15, 10	10, 3	8, 5	8	1/4	2754	176.0	2.80e–11
7	$10^{-7}\sim 10^{-5}$	10, 15	5, 13	4, 6	10	1/3	3560	82.70	3.10e–11
8	$10^{-7}\sim 10^{-5}$	25, 25	20, 20	7, 7	15	1/4	6881	415.6	7.27e–12
9	$10^{-8}\sim 10^{-6}$	35, 35	30, 30	6, 6	20	1/6	5909	658.1	1.44e–14
10	$10^{-8}\sim 10^{-6}$	45, 45	40, 40	6, 6	5	1/5	4327	521.1	7.61e–14
11	$10^{-8}\sim 10^{-6}$	75, 70	60, 60	7, 7	4	1/10	7832	1860	3.78e–14
12	$10^{-8}\sim 10^{-6}$	85, 85	80, 80	3, 3	5	1/10	4757	407.4	3.64e–14
13	$10^{-9}\sim 10^{-7}$	25, 25	20, 20	5, 5	77	1/8	11073	6868	9.84e–17

Table 1: Algorithm performance on benchmarks

Windows for *Digits*:=15. In Table 1 we exhibit the performance of our algorithm for recovering multivariate rational functions from a black box that returns noisy values with outlier errors. For each example, we construct two relatively prime polynomials with random integer coefficients in the range $-5 \le c \le 5$. Here *Random Noise* denotes the relative noise in this range randomly added to the black box of f/g; $\bar{d}_f \ge \deg(f)$ and $\bar{d}_g \ge \deg(g)$ denote the degree bound of the numerator and denominator, respectively; t_f and t_g denote the number of terms of the numerator and denominator, respectively; n denotes the number of variables of the rational functions; N denotes the number of the black box probes needed to interpolate the approximate multivariate rational function; $1/q$ is the error rate of the outlier error; *Rel. Error* is the relative error, namely $(\|c\tilde{f} - f\|_2^2 + \|c\tilde{g} - g\|_2^2)/(\|f\|_2^2 + \|g\|_2^2)$, where $\tilde{f}/\tilde{g}$ is the fraction computed by our algorithm and c is optimally chosen to minimize the error. For each example, the outlier error is the relative error of the evaluation, which is in the range of $0.01 \times [100, 200]$. In Table 1 and Table 2, the upper bound E is chosen as $E = \lfloor(\bar{d}_f + \bar{d}_g + 1)/(q-2)\rfloor$, for the given error rate $1/q$ and the degree bounds $\bar{d}_f$, $\bar{d}_g$. In the step of getting the evaluations for applying univariate dense Cauchy interpolation algorithm, i.e., Step 2(a.2) of our algorithm, we choose k error locations randomly, and then add the outlier errors. Note that the actual count of errors k is also chosen randomly in the range $0 \le k \le E$. Running times serve to give a rough idea on the efficiency, and are for SONY VAIO laptops with 8GB of memory and 2.67GHz Intel i7 processors.

E x	Random Noise	Rel. Outlier Error Θ	deg(f), deg(g)	t_f, t_g	n	1/q	N	time secs	Rel. Error
1	$10^{-6}\sim 10^{-4}$	1∼2	3, 3	2, 3	2	1/4	203	2.371	2.83e–9
2	$10^{-7}\sim 10^{-5}$	0.1∼0.2	10, 10	3, 5	2	1/8	558	14.80	3.16e–12
3	$10^{-7}\sim 10^{-5}$	0.001∼0.002	10, 3	4, 3	3	1/7	822	15.09	2.92e–12
4	$10^{-7}\sim 10^{-5}$	0.01∼0.02	5, 5	4, 4	5	1/4	960	10.58	4.72e–11
5	$10^{-8}\sim 10^{-6}$	0.001∼0.002	10, 10	5, 4	7	1/10	1002	26.93	1.36e–13
6	$10^{-8}\sim 10^{-6}$	0.1∼0.2	5, 8	1, 3	10	1/5	2010	37.58	2.25e–12
7	$10^{-6}\sim 10^{-4}$	0.01∼0.02	10, 15	3, 3	4	1/10	1874	88.09	8.12e–12
8	$10^{-7}\sim 10^{-5}$	0.01∼0.02	10, 10	3, 2	15	1/5	2786	44.66	6.32e–12
9	$10^{-8}\sim 10^{-6}$	0.01∼0.02	8, 8	4, 3	30	1/5	4798	206.0	1.58e–13
10	$10^{-9}\sim 10^{-7}$	0.01∼0.02	15, 15	3, 3	50	1/4	14717	2080	1.90e–16

Table 2: Alg. performance on benchmarks (small outliers)

In Table 2 we give tests with small outlier errors. Here *Outlier Error* denotes the relative outlier error Θ, which is randomly selected in the given range. $\bar{d}_f$, $\bar{d}_g$ are chosen by $\bar{d}_f = \deg(f) + 5, \bar{d}_g = \deg(g) + 5$.

6. REFERENCES

[1] Arnold, A., Giesbrecht, M., and Roche, D. S. Faster sparse interpolation of straight-line programs. In *CASC* (2013), V. P. Gerdt, W. Koepf, E. W. Mayr, and E. V. Vorozhtsov, Eds., vol. 8136 of *LNCS*, Springer, pp. 61–74.

[2] Ben-Or, M., and Tiwari, P. A deterministic algorithm for sparse multivariate polynomial interpolation. In *Proc. Twentieth STOC* (New York, N.Y., 1988), ACM Press, pp. 301–309.

[3] Blahut, R. E. A universal Reed-Solomon decoder. *IBM J. Res. Develop. 18*, 2 (Mar. 1984), 943–959.

[4] Candès, E. J., and Tao, T. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inform. Theory 52* (2004), 5406–5425.

[5] Comer, M. T., Kaltofen, E. L., and Pernet, C. Sparse polynomial interpolation and Berlekamp/Massey algorithms that correct outlier errors in input values. In *Proc. ISSAC 2012* (New York, N. Y., July 2012), J. van der Hoeven and M. van Hoeij, Eds., ACM, pp. 138–145.

[6] Cuyt, A., and Lee, W. Sparse interpolation of multivariate rational functions. *TCS 412* (2011), 1445–1456.

[7] de Kleine, J., Monagan, M., and Wittkopf, A. Algorithms for the non-monic case of the sparse modular GCD algorithm. In *Proc. ISSAC 2005* (New York, N. Y., 2005), M. Kauers, Ed., ACM Press, pp. 124–131.

[8] Garg, S., and Schost, Éric. Interpolation of polynomials given by straight-line programs. *TCS 410*, 27-29 (2009), 2659–2662.

[9] Giesbrecht, M., Labahn, G., and Lee, W. Symbolic-numeric sparse interpolation of multivariate polynomials. *JSC 44* (2009), 943–959.

[10] Kaltofen, E. Factorization of polynomials given by straight-line programs. In *Randomness and Computation*, S. Micali, Ed. JAI Press Inc., 1989, pp. 375–412.

[11] Kaltofen, E. Fifteen years after DSC and WLSS2 What parallel computations I do today [Invited lecture at PASCO 2010]. In *Proc. PASCO 2010* (New York, N. Y., July 2010), M. Moreno Maza and J.-L. Roch, Eds., ACM, pp. 10–17.

[12] Kaltofen, E., Lakshman Y. N., and Wiley, J. M. Modular rational sparse multivariate polynomial interpolation. In *Proc. 1990 Internat. Symp. Symbolic Algebraic Comput. (ISSAC'90)* (1990), S. Watanabe and M. Nagata, Eds., ACM Press, pp. 135–139.

[13] Kaltofen, E., and Lee, W. Early termination in sparse interpolation algorithms. *JSC 36*, 3–4 (2003), 365–400.

[14] Kaltofen, E., and Nehring, M. Supersparse black box rational function interpolation. In *Proc. 2011 Internat. Symp. Symbolic Algebraic Comput. ISSAC 2011* (New York, N. Y., June 2011), A. Leykin, Ed., ACM, pp. 177–185.

[15] Kaltofen, E., and Pernet, C. Cauchy interpolation with errors in the values. Submitted manuscript, 13 pages, Dec. 2013.

[16] Kaltofen, E., and Yang, Z. On exact and approximate interpolation of sparse rational functions. In *Proc. ISSAC 2007* (New York, N. Y., 2007), C. W. Brown, Ed., ACM Press, pp. 203–210.

[17] Kaltofen, E., and Yang, Z. Sparse multivariate function recovery from values with noise and outlier errors. In *Proc. ISSAC 2013* (New York, N. Y., 2013), M. Kauers, Ed., ACM, pp. 219–226.

[18] Kaltofen, E., Yang, Z., and Zhi, L. On probabilistic analysis of randomization in hybrid symbolic-numeric algorithms. In *Proc. SNC 2007* (New York, N. Y., 2007), J. Verschelde and S. M. Watt, Eds., ACM Press, pp. 11–17.

[19] Mahoney, M. W. Randomized algorithms for matrices and data. URL: http://arxiv.org/abs/1104.5557, 2011.

[20] Olshevsky, V., and Shokrollahi, M. A. A displacement approach to decoding algebraic codes. In *Algorithms for Structured Matrices: Theory and Applications*. AMS, Providence, RI, 2003, pp. 265–292. Contemp. Math., vol. 323.

[21] Prony, R. Essai expérimental et analytique sur les lois de la Dilatabilité de fluides élastiques etc. *J. de l'École Polytechnique 1* (Floréal et Prairial III (1795)), 24–76.

[22] Saraf, S., and Yekhanin, S. Noisy interpolation of sparse polynomials, and applications. In *Proc. 26th Annual IEEE Conf. Comp. Complexity* (2011), IEEE Comp. Soc., pp. 86–92.

[23] Welch, L. R., and Berlekamp, E. R. Error correction of algebraic block codes. US Patent 4,633,470, 1986. Filed 1983.

[24] Zippel, R. Interpolating polynomials from their values. *JSC 9*, 3 (1990), 375–403.

[25] Zippel, R. E. *Probabilistic algorithms for sparse polynomials.* PhD thesis, MIT, Cambridge, USA, Sept. 1979.