Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

A Bayesian Model for Root Computation

H.K. Pikkarainen, J. Schicho

RICAM-Report 2008-31

A Bayesian Model for Root Computation

Hanna K. Pikkarainen* Josef Schicho*

Abstract

A univariate polynomial over the real or the complex numbers is given approximately. We present a Bayesian method for the computation of the posterior probabilities of different multiplicity patterns. The method is based on interpreting the root computation problem as an inverse problem which is then treated in the Bayesian framework. The performance of the method is illustrated by several numerical examples.

Keywords: Root computation, Univariate polynomials, Multiplicity patterns, Bayesian inversion theory.

Mathematics Subject Classification (2000): 26C10, 12Y05

1 Introduction

The problem of dealing with multiplicities in the root computation for a numerically given univariate polynomial is relevant for various applications in algebraic computation. There are applications where multiple roots are either expected, or at least there is a deliberate desire for not to overlook the possibility of multiple roots. For instance, one may be interested in the singularities of a plane algebraic curve; such singularities cause multiple roots of the discriminant (a univariate polynomial).

In the fast algorithms described in [3, 11, 10] (see also [7]), the algebraic concept of the multiplicity is replaced by the numerical concept of a "root

*Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria; first-name.lastname@oeaw.ac.at

cluster". In general, a root cluster may contain several distinct roots; this may be detected if the precision is increased, but it is not always possible due to accuracy limitations of the given data.

The approach in [12] is to fix the multiplicity structure, restricting the problem to a "pejorative manifold", and then to solve a nonlinear least squares problem. It turns out that this computation is numerically robust; so, once we know the multiplicities, the numerical values of the roots may be computed efficiently and accurately. An essential step is to estimate the multiplicity structure. This is done via approximate GCD computation, based on results of [5, 4]. The problem is reduced to estimating the rank of a numerically given matrix; this estimation is based on heuristics and is confirmed by the success of examples. Even if the estimate was not "correct" (meaning that the original polynomial before the perturbation had a different multiplicity structure), the computed estimates make sense from a numerical point of view.

In this paper, we deal with the problem of inferring the multiplicity structure from the given data using a statistical model. The root computation problem is interpreted as an inverse problem and the corresponding inverse problem is solved by utilizing the Bayesian inversion theory. The Bayesian framework allows to compute the posterior distributions of the roots, and therefore various estimates for them (e.g., the conditional mean estimate). Since the Bayesian approach to inverse problems is known to be computationally demanding, good estimates can be computed more efficiently, for instance as in [12]. On the other hand, the Bayesian framework has the advantage that it allows to obtain quite precise qualitative information about the roots. Under the assumption that the coefficients of the polynomial are affected by an additive noise with a known Gaussian probability distribution, and under the assumption of a non-informative prior distribution, we compute the posterior probability of all multiplicity patterns.

Guideline for the reader: in Section 2, we give a general overview on the Bayesian approach to inverse problems and also prove a theoretical result which is not yet available in the literature (Theorem 2.2). Section 3 describes the application of the Bayesian method to the problem of computing multiplicity information for univariate polynomials. Section 4 shows by concrete examples that (and how) the posterior probabilities of different multiplicity patterns can be computed effectively, and illustrates various phenomena concerning the posterior probabilities of multiple roots. Conclusions are given in Section 5.

2 Bayesian approach to inverse problems

Inverse problems are encountered typically in situations where one makes indirect observations of a quantity of interest. Root computation can be interpreted as an inverse problem. One is interested in quantitative and qualitative properties of the roots of an univariate polynomial but can only observe the coefficients of the polynomial.

The interdependency between the unknown quantity and the observed quantity in an inverse problem is described through a mathematical model. In root computation, there is a known nonlinear mapping between the roots and the coefficients of an univariate polynomial, which depends on the degree of the polynomial.

Usually, the data in an inverse problem, i. e., indirect observations are noisy in a sense that the value of the observed quantity is not known exactly. In root computation, the coefficient of a polynomial are given only numerically and may contain some errors.

There are several numerical methods for computing an estimate for the quantity of interest in the presence of noisy indirect observations. Since we are mainly interested in the qualitative properties of the roots, in this paper we concentrate on the Bayesian approach to inverse problems and use it for solving the root compution problem. In this section, we summarise the main ideas of the Bayesian inversion theory. A comprehensive introduction into the topic can be found in [8].

In the Bayesian framework all quantities included in the model of an inverse problem are treated as random variables. The probabilities appearing in the Bayesian approach need not correspond to frequencies of random events but they are also used to describe the confidence or the degree of belief that one has into a particular initial guess.

All information available before observations about the quantity of interest is coded into a probability distribution, the so-called *prior distribution*. Even though the quantity of interest would be assumed to be deterministic, it is modelled by a random variable whose distribution is the prior distribution.

In the Bayesian framework the solution of an inverse problem is the *posterior distribution*, i.e., the conditional distribution of the random variable of interest with respect to the observed data. When the joint distribution of the random variables describing the unknown and the observations is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^n$, the Bayesian inversion theory is based on the Bayes formula. The Bayes formula describes how the prior information and observations have to be combined to give the posterior distribution; by this formula the posterior probability density is proportional to the product of the prior probability density and the *likelihood function* which is given by the model for the indirect observations (see, e.g., [8]).

Usually only real inverse problems are considered in the Bayesian framework. We are interested in roots of real and complex univariate polynomials. Since $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ are isomorphic, the Bayesian inversion theory can also be applied to complex inverse problems.

In root computation it is desirable that the probability of the appearance of multiple roots is positive. Hence the prior distribution cannot be absolutely continuous with respect to the Lebesgue measure since the Lebesgue measure of lower-dimensional subspaces of $\mathbb{C}^n$ is zero. Thus we cannot apply the basic form of the Bayes formula. A modification of the Bayes formula suitable for the root computation problem is presented in this section in a general framework.

At first, we give the definition of the conditional distribution. We denote random variables by capital letters and their realizations by lower case letters. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(G, \mathcal{G})$ and $(H, \mathcal{H})$ be measurable spaces, and $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (G, \mathcal{G})$ and $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to (H, \mathcal{H})$ be random variables. We denote the distribution of Y by μ_Y. A *conditional distribution of X given Y* is a function $\mu_{X|Y}(\cdot\,|\,\cdot) : \mathcal{G} \times H \to [0, 1]$ satisfying the following conditions:

(i) $\mu_{X|Y}(\cdot\,|\,y)$ is a probability measure in $(G, \mathcal{G})$ for μ_Y-almost all $y \in H$,

(ii) $\mu_{X|Y}(A\,|\,\cdot)$ is $\mathcal{H}$-measurable for all $A \in \mathcal{G}$, and

(iii) for all $A \in \mathcal{G}$ and $B \in \mathcal{H}$

$$\int_B \mu_{X|Y}(A\,|\,y)\, \mathrm{d}\mu_Y(y) = \mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)).$$

The existence of the conditional distribution is guaranteed if G is a Polish space (i.e., a separable completely metrizable topological space).

Theorem 2.1. [6, Section 10.2] *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(G, \mathcal{G})$ and $(H, \mathcal{H})$ be measurable spaces and $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (G, \mathcal{G})$ and $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to (H, \mathcal{H})$ be random variables. If G is a Polish space, a conditional distribution $\mu_{X|Y}$ of X given Y exists. In addition, if $\tilde{\mu}_{X|Y}$ also satisfies the definition of a conditional distribution of X given Y, the measures $\mu_{X|Y}(\cdot\,|\,y)$ and $\tilde{\mu}_{X|Y}(\cdot\,|\,y)$ are equal for μ_Y-almost all $y \in H$.*

Let X and Y be random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to measurable spaces $(G, \mathcal{G})$ and $(H, \mathcal{H})$, respectively. We suppose that the random variable X is unobservable and of our primary interest and Y is directly observable. We call X the *unknown*, Y the *observation* and its realization y_{data} in the actual observation process the *data*. We assume that we have a model for the observations with additive noise

$$Y = F(X) + E \tag{1}$$

where $F : (G, \mathcal{G}) \to (H, \mathcal{H})$ is a measurable function and $E : (\Omega, \mathcal{F}, \mathbb{P}) \to (H, \mathcal{H})$ is a random variable. The inverse problem is to find the conditional distribution of the unknown X given the data $Y = y_{\text{data}}$ when the observation Y satisfies the additive noise model (1). Let us denote the distributions of X, Y, and E by μ_X, μ_Y, and μ_E, respectively. The probability measure μ_X is the prior distribution of the unknown X.

Since we want to emphasis that the randomness of the prior distribution and the noise E arises from different sources, we assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a product space $(\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \times \mathcal{F}_2}, \mathbb{P}_1 \times \mathbb{P}_2)$ and $X(\omega) = X(\omega_1)$ and $E(\omega) = E(\omega_2)$ for all $\omega = (\omega_1, \omega_2) \in \Omega$. Then X and E are independent.

If G is a Polish space, the conditional distribution of X given Y exists. Hence the solution of the inverse problem (1) exists and the posterior distribution μ_{post} of X given the data $Y = y_{\text{data}}$ is

$$\mu_{\text{post}}(A) = \mu_{X|Y}(A \,|\, y_{\text{data}}), \quad A \in \mathcal{G},$$

for μ_Y-almost all y_{data}. When the noise distribution μ_E behaves well enough, the posterior distribution μ_{post} can be calculated by the formula in the following theorem.

Theorem 2.2. *Let $X : (\Omega_1, \mathcal{F}_1, \mathbb{P}_1) \to (G, \mathcal{G})$ and $E : (\Omega_2, \mathcal{F}_2, \mathbb{P}_2) \to (H, \mathcal{H})$ be random variables and $F : (G, \mathcal{G}) \to (H, \mathcal{H})$ be a measurable function. Assume that Y satisfies the additive noise model* (1) *in $(\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \times \mathcal{F}_2}, \mathbb{P}_1 \times \mathbb{P}_2)$. For all $h \in H$ denote by μ_{E+h} the distribution of $E + h$. Assume that μ_{E+h} is absolutely continuous with respect to μ_E for all $h \in H$, the Radon-Nikodym derivative is positive, i. e.,*

$$\frac{\mathrm{d}\mu_{E+h}}{\mathrm{d}\mu_E}(y) > 0$$

μ_E-almost all $y \in H$, and the function $(x,y) \mapsto \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)$ is measurable in $(G \times H, \mathcal{G} \times \mathcal{H})$. If G is a Polish space,

$$\mu_{X|Y}(A \,|\, y) = \frac{\int_A \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x)}{\int_G \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x)} \tag{2}$$

for μ_Y-almost all $y \in H$ and all $A \in \mathcal{G}$.

Proof. Let $A \in \mathcal{G}$ and $B \in \mathcal{H}$. By using the additive noise model (1) and the Fubini theorem,

$$\begin{aligned}
&\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) \\
&= \int_{\Omega_1 \times \Omega_2} \chi_A(X(\omega_1))\chi_B(F(X(\omega_1)) + E(\omega_2))\ \mathbb{P}_1 \times \mathbb{P}_2(\mathrm{d}\omega_1, \mathrm{d}\omega_2) \\
&= \int_{\Omega_1} \left[\int_{\Omega_2} \chi_B(F(X(\omega_1)) + E(\omega_2))\ \mathbb{P}_2(\mathrm{d}\omega_2)\right] \chi_A(X(\omega_1))\ \mathbb{P}_1(\mathrm{d}\omega_1) \\
&= \int_{\Omega_1} \left[\int_B \mu_{E+F(X(\omega_1))}(\mathrm{d}y)\right] \chi_A(X(\omega_1))\ \mathbb{P}_1(\mathrm{d}\omega_1).
\end{aligned}$$

Since μ_{E+h} is absolutely continuous with respect to μ_E for all $h \in H$,

$$\begin{aligned}
&\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) \\
&= \int_{\Omega_1} \left[\int_B \frac{\mathrm{d}\mu_{E+F(X(\omega_1))}}{\mathrm{d}\mu_E}(y)\ \mu_E(\mathrm{d}y)\right] \chi_A(X(\omega_1))\ \mathbb{P}_1(\mathrm{d}\omega_1) \\
&= \int_{\Omega_1} \left[\int_{\Omega_2} \chi_B(E(\omega_2))\frac{\mathrm{d}\mu_{E+F(X(\omega_1))}}{\mathrm{d}\mu_E}(E(\omega_2))\ \mathbb{P}_2(\mathrm{d}\omega_2)\right] \chi_A(X(\omega_1))\ \mathbb{P}_1(\mathrm{d}\omega_1) \\
&= \int_{\Omega_1 \times \Omega_2} \chi_A(X(\omega_1))\chi_B(E(\omega_2))\frac{\mathrm{d}\mu_{E+F(X(\omega_1))}}{\mathrm{d}\mu_E}(E(\omega_2))\ \mathbb{P}_1 \times \mathbb{P}_2(\mathrm{d}\omega_1, \mathrm{d}\omega_2).
\end{aligned}$$

According to the assumptions the random variables X and E are independent. Hence by the Fubini theorem,

$$\begin{aligned}
\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) &= \int_{A \times B} \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X \times \mu_E(\mathrm{d}x, \mathrm{d}y) \\
&= \int_B \left[\int_A \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x)\right]\ \mu_E(\mathrm{d}y).
\end{aligned}$$

Because

$$\mu_Y(B) = \mathbb{P}(X^{-1}(G) \cap Y^{-1}(B)) = \int_B \left[\int_G \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x) \right]\ \mu_E(\mathrm{d}y),$$

the distribution of Y is absolutely continuous with respect to the distribution of E and

$$\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_E}(y) = \int_G \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x)$$

for μ_E-almost all $y \in H$. By the assumptions $\frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y) > 0$ for all $x \in G$ and μ_E-almost all $y \in H$ and therefore $\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_E}(y) > 0$ for μ_E-almost all $y \in H$. Thus $\mu_Y(B) = 0$ if and only if $\mu_E(B) = 0$ for all $B \in \mathcal{H}$. Hence μ_E is absolutely continuous with respect to μ_Y and

$$\frac{\mathrm{d}\mu_E}{\mathrm{d}\mu_Y}(y) = \left(\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_E}(y) \right)^{-1}$$

for μ_Y-almost all $y \in H$. Therefore

$$\begin{aligned}
&\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) \\
&= \int_B \left[\int_A \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x) \right] \frac{\mathrm{d}\mu_E}{\mathrm{d}\mu_Y}(y)\ \mu_Y(\mathrm{d}y) \\
&= \int_B \left[\int_A \frac{\mathrm{d}\mu_{E+F(x)}}{\mathrm{d}\mu_E}(y)\ \mu_X(\mathrm{d}x) \right] \left(\frac{\mathrm{d}\mu_Y}{\mathrm{d}\mu_E}(y) \right)^{-1}\ \mu_Y(\mathrm{d}y).
\end{aligned}$$

The claim is proven since by the assumptions the right hand side of (2) fulfills also the conditions (i) and (ii) of the conditional distribution. □

Theorem 2.2 can be seen as a generalized version of the Bayes formula. The proof is a modification of the proof of [9, Theorem 4], which gives a formula to calculate conditional expectations in the case of additive normal noise in real separable Banach spaces.

The assumptions of Theorem 2.2 requires that the support of the noise distribution is unbounded. For example, normal noise distributions in $\mathbb{C}^n$ and $\mathbb{R}^n$ satisfy the assumptions of Theorem 2.2 (see Section 3).

Even though the posterior distribution is the solution of an inverse problem in the Bayesian framework, often point and spread estimates are calculated to give a more concrete view of the solution. The most common point estimate is the *conditional mean* (CM) estimate

$$x_{\mathrm{CM}} = \int_G x \, \mu_{\mathrm{post}}(\mathrm{d}x)$$

provided that the integral exists.

If G is a Banach space, a typical spread estimate is the *conditional covariance operator* $\mathrm{cov}(X \mid y_{\mathrm{data}}) : G^* \to G$ defined as

$$\mathrm{cov}(X \mid y_{\mathrm{data}})x^* = \int_G \langle x^*, x - x_{\mathrm{CM}} \rangle (x - x_{\mathrm{CM}}) \, \mu_{\mathrm{post}}(\mathrm{d}x)$$

for all $x^* \in G^*$ if the integral exists. The conditional mean and the conditional covariance operator are also called the posterior mean and the posterior covariance operator, respectively, since there are the mean and the covariance operator of the posterior distribution if they exist.

In our application, mainly qualitative information about the solution of an inverse problem is wanted. E. g., in root computation the probability of the existence of a certain root configuration may be of interest. Let A be the set that describe the desired qualitative property. Then the posterior probability of A is

$$\mu_{\mathrm{post}}(A) = \int_A \mu_{\mathrm{post}}(\mathrm{d}x).$$

3 Polynomials of degree n

In this section the root computation problem is treated in the Bayesian framework and the posterior probabilities of various multiplicity patterns are calculated. The cases of complex and real coefficients are studied separately since for purely real univariate polynomials the additional information about the roots, i. e., roots are either real or complex conjugate pairs, should be taken into account in the prior distribution.

3.1 Complex coefficients

We assume that the complex coefficients a_i, $i = 1, \ldots, n$, of the univariate polynomial $P(z) = z^n + a_1 z^{n-1} + \ldots + a_{n-1}z + a_n$ are given. We want to calculate the complex roots z_i, $i = 1, \ldots, n$, of the polynomial. For a polynomial of degree n, the known connection between the coefficients and the roots is

$$a_i = (-1)^i \sum_{1 \le j_1 < j_2 < \ldots < j_i \le n} z_{j_1} z_{j_2} \ldots z_{j_i} \tag{3}$$

for all $i = 1, \ldots, n$. The root computation problem can be identified with the inverse problem to find an estimate for the roots $(z_1, z_2, \ldots, z_n) \in \mathbb{C}^n$ for given coefficients $(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n$. Formula (3) gives the required mapping from the unknown roots to the known coefficients.

However, the order of the roots does not play a role in the root computation problem. Hence we want to identify all permutations of roots. We define the equivalence relation $\sim$ in $\mathbb{C}^n$ by $z \sim w$ if and only if $z = (w_{i_1}, w_{i_2}, \ldots, w_{i_n})$ where $(i_1, i_2, \ldots, i_n) = p(1, 2, \ldots, n)$ for some permutation $p \in S_n$. We denote the equivalence class of z by $[z]$. In this paper the solution of the root computation problem is an equivalence class in the quotient space $\mathbb{C}^n/\sim$, not a single vector in $\mathbb{C}^n$.

We define the function $F : \mathbb{C}^n/\sim \to \mathbb{C}^n$ by

$$F([z]) = \begin{pmatrix} -(z_1 + z_2 + \ldots + z_n) \\ \vdots \\ (-1)^i \sum_{1 \le j_1 < j_2 < \ldots < j_i \le n} z_{j_1} z_{j_2} \ldots z_{j_i} \\ \vdots \\ (-1)^n z_1 z_2 \ldots z_n \end{pmatrix}.$$

Note that the function F is well-defined. We assume that $\mathbb{C}^n/\sim$ is equipped with the quotient topology, i. e., $U \subset \mathbb{C}^n/\sim$ is open if and only if $\{z \in \mathbb{C}^n : [z] \in U\}$ is open in $\mathbb{C}^n$. Then F is measurable with respect to the Borel σ-algebras $\mathcal{B}(\mathbb{C}^n/\sim)$ and $\mathcal{B}(\mathbb{C}^n)$ of $\mathbb{C}^n/\sim$ and $\mathbb{C}^n$, respectively.

The coefficients $a = (a_1, a_2, \ldots, a_n)$ are known only approximately. We want to utilize the Bayesian inversion theory for solving the root computation problem. Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces. We assume that the function values of F are perturbed by an additive noise E. Then the inverse problem can be modeled by

$$A = F([Z]) + E$$

where $A : (\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \times \mathcal{F}_2}, \mathbb{P}_1 \times \mathbb{P}_2) \to \mathbb{C}^n$, $[Z] : (\Omega_1, \mathcal{F}_1, \mathbb{P}_1) \to \mathbb{C}^n/\sim$ and $E : (\Omega_2, \mathcal{F}_2, \mathbb{P}_2) \to \mathbb{C}^n$ are random variables. Here, the notation $[Z]$ is used as a name for a random variables in $\mathbb{C}^n/\!\sim$ so that its realization $[z]$ inherits the notation usual in $\mathbb{C}^n/\!\sim$.

The distribution μ_E of the noise is assumed to be the complex normal distribution. As usual the mean of the noise distribution is supposed to be zero. Since by the function F the unit of the coefficient a_i is the i^{th} power of the unit of the roots, the covariance matrix of the noise distribution is assumed to be $\varepsilon^2 J$ where $\varepsilon > 0$ and $J \in \mathbb{C}^{n\times n}$ is the diagonal matrix $J = \operatorname{diag}(N, N^2, \ldots, N^n)$ for some $N > 0$. Then the distribution of the noise has the density function

$$\pi_E(w) = \frac{1}{\pi^n \varepsilon^{2n} N^{n(n+1)/2}} \exp\left(-\frac{w^* J^{-1} w}{\varepsilon^2}\right), \quad w \in \mathbb{C}^n,$$

with respect to the Lebesgue measure in $\mathbb{C}^n$ (see, e. g., [1, Chapter 2]).

The distribution μ_{E+h} is absolutely continuous with respect to μ_E for all $h \in \mathbb{C}^n$ and the Radon-Nikodym derivative of μ_{E+h} with respect to μ_E is

$$\frac{\mathrm{d}\mu_{E+h}}{\mathrm{d}\mu_E}(w) = \exp\left(-\frac{h^* J^{-1} h - 2\mathcal{R}e(w^* J^{-1} h)}{\varepsilon^2}\right)$$

for all $w \in \mathbb{C}^n$. Hence the complex normal distribution fulfills the assumptions of Theorem 2.2 for the noise distribution.

Since $\mathbb{C}^n/\!\sim$ is homeomorphic to $\mathbb{C}^n$, by the definition it is a Polish space. Thus by Theorem 2.2 the posterior distribution μ_{post} of the roots $[Z]$ given the noisy coefficients $a = (a_1, a_2, \ldots, a_n)$ is

$$\mu_{\text{post}}(U) = \frac{\int_U \exp\left(-\frac{(F([z])-a)^* J^{-1}(F([z])-a)}{\varepsilon^2}\right) \mu_{\text{pr}}(\mathrm{d}[z])}{\int_{\mathbb{C}^n/\sim} \exp\left(-\frac{(F([z])-a)^* J^{-1}(F([z])-a)}{\varepsilon^2}\right) \mu_{\text{pr}}(\mathrm{d}[z])}$$

for almost all $a \in \mathbb{C}^n$ where $U \in \mathcal{B}(\mathbb{C}^n/\!\sim)$ and μ_{pr} is a prior distribution of the roots in $\mathbb{C}^n/\!\sim$.

In this paper we assume that all information about the roots of a complex univariate polynomial is given by the coefficients. Hence we should use a non-informative prior distribution in $\mathbb{C}^n/\!\sim$, e. g., to lift an uniform distribution on a disc with radius M for all roots from $\mathbb{C}^n$ to $\mathbb{C}^n/\!\sim$. Since all possible multiplicities of n roots should be taking into account, we use the prior distribution μ_{pr} which is a combination of lifted uniform distributions

$$\mu_{\text{pr}}(U) = \sum_{k=1}^{n} \sum_{\substack{i_1+i_2+\ldots+i_k=n \\ 1\le i_1\le i_2\le\ldots\le i_k\le n}} \frac{\alpha_{i_1,i_2,\ldots,i_k}}{\pi^k M^{2k}} \int_{\{(w_1,\ldots,w_k)\in\mathbb{C}^k : [w_{i_1,i_2,\ldots,i_k}]\in U\}} \chi_{B_0(M)^k}(w_1,\ldots,w_k)\, \mathrm{d}(w_1,\ldots,w_k)$$

where $U \in \mathcal{B}(\mathbb{C}^n/\sim)$, $B_0(M)$ is the disc in $\mathbb{C}$ with centre 0 and radius $M > 0$,

$$w_{i_1,i_2,\ldots,i_k} := (\underbrace{w_1,\ldots,w_1}_{i_1},\underbrace{w_2,\ldots,w_2}_{i_2},\ldots,\underbrace{w_k,\ldots,w_k}_{i_k}) \in \mathbb{C}^n$$

for all $w = (w_1, w_2, \ldots, w_k) \in \mathbb{C}^k$, and $\alpha_{i_1,i_2,\ldots,i_k}$ are constants on the interval $[0, 1]$ satisfying

$$\sum_{k=1}^{n} \sum_{\substack{i_1+i_2+\ldots+i_k=n \\ 1\le i_1\le 1_2\le\ldots\le i_k\le n}} \alpha_{i_1,i_2,\ldots,i_k} = 1.$$

The constant $\alpha_{i_1,i_2,\ldots,i_k}$ is the prior probability of the appearance of k separate roots with multiplicities i_1, i_2, …, and i_k for all $k = 1, \ldots, n$.

Let us use the notation $\|w\|_J^2 = w^* J^{-1} w$ for all $w \in \mathbb{C}^n$. With the above choice of the prior distribution, we need to calculate integrals of the form

$$\int_{\{w\in B_0(M)^k : [w_{i_1,i_2,\ldots,i_k}]\in U\}} \exp\left(-\frac{1}{\varepsilon^2}\|F([w_{i_1,i_2,\ldots,i_k}]) - a\|_J^2 \right) \mathrm{d}w$$

for all $(i_1, i_2, \ldots, i_k)$ such that $1 \le i_1 \le i_2 \le \ldots \le i_k \le n$ and $i_1+i_2+\ldots+i_k = n$ where $U \in \mathcal{B}(\mathbb{C}^n/\sim)$ and $k = 1, \ldots, n$ if we want to evaluate the posterior probability of a given measurable set.

A possible question of interest is what is the posterior probability of the appearance of at least one multiple root. Let $U := \{[z] \in \mathbb{C}^n/\sim : z_i = z_j \text{ for some } i \ne j\}$. The probability we are interested in is $\mu_{\text{post}}(U)$. Let us denote

$$p_{i_1,i_2,\ldots,i_k} = \frac{\alpha_{i_1,i_2,\ldots,i_k}}{\pi^k M^{2k}} \int_{B_0(M)^k} \exp\left(-\frac{1}{\varepsilon^2}\|F([w_{i_1,i_2,\ldots,i_k}]) - a\|_J^2 \right) \mathrm{d}w$$

for all $(i_1, i_2, \ldots, i_k)$ such that $1 \le i_1 \le i_2 \le \ldots \le i_k \le n$ and $i_1+i_2+\ldots+i_k = n$ where $k = 1, \ldots, n$. Then

$$\mu_{\text{post}}(U) = 1 - \mu_{\text{post}}(\mathbb{C}^n/\sim \setminus U) = 1 - \frac{p_{1,1,\ldots,1}}{\sum_{k=1}^{n} \sum_{\substack{i_1+i_2+\ldots+i_k=n \\ 1\le i_1\le i_2\le\ldots\le i_k\le n}} p_{i_1,i_2,\ldots,i_k}}.$$

Remark 3.1. For calculating the posterior probabilities of different root configurations for complex univariate polynomials the noise and the prior distributions used in this subsection are not the only possible ones. The method requires only that the noise distribution fulfills the assumptions of Theorem 2.2. For the prior distribution the uniform distributions can be replaced by other distributions suitable from the application point of view.

3.2 Real coefficients

If the coefficients a_i, $i = 1, \ldots, n$, of the univariate polynomial $P(z) = z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n$ are real, the roots of the polynomial are either real or complex conjugate pairs. The method presented in the previous subsection could also be applied to the real case but actually the additional information about the roots of real univariate polynomials should be incorporated in the prior distribution. In this subsection we modify the method of the previous subsection to be more suitable for real univariate polynomials.

Let $R \subset \mathbb{C}^n$ be the set of all possible roots of real univariate polynomials, i. e., $R = \{z \in \mathbb{C}^n : z_i \in \mathbb{R} \vee (\exists\, w \in \mathbb{C} \setminus \mathbb{R} \text{ and } k \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor\} : z_{i_1} = \ldots = z_{i_k} = w \text{ and } z_{j_1} = \ldots = z_{j_k} = \bar{w})\}$. Since the set R is closed in $\mathbb{C}^n$, also $R/{\sim}$ is closed in $\mathbb{C}^n/{\sim}$ and hence the function F restricted to $R/{\sim}$ is measurable from $(R/{\sim}, \mathcal{B}(R/{\sim}))$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The inverse problem in the Bayesian framework related to root computation of real univariate polynomial is

$$A = F([Z]) + E \tag{4}$$

where $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ are probability spaces, and $A : (\Omega_1 \times \Omega_2, \overline{\mathcal{F}_1 \times \mathcal{F}_2}, \mathbb{P}_1 \times \mathbb{P}_2) \to \mathbb{R}^n$, $[Z] : (\Omega_1, \mathcal{F}_1, \mathbb{P}_1) \to R/{\sim}$ and $E : (\Omega_2, \mathcal{F}_2, \mathbb{P}_2) \to \mathbb{R}^n$ are random variables.

Since the coefficients a_i, $i = 1, \ldots, n$, are real, we can utilize as a noise distribution in the additive noise model (4) a real probability distribution. Similarly as in the complex case, we assume that the noise distribution μ_E is the normal distribution with the mean 0 and the covariance matrix $\varepsilon^2 J$ where $\varepsilon > 0$ and $J \in \mathbb{R}^{n \times n}$ is the diagonal matrix $J = \operatorname{diag}(N, N^2, \ldots, N^n)$ for some $N > 0$. Then the distribution of the noise has the density function

$$\pi_E(y) = \frac{1}{\sqrt{(2\pi)^n \varepsilon^{2n} N^{n(n+1)/2}}} \exp\left(-\frac{y^T J^{-1} y}{2\varepsilon^2}\right), \quad y \in \mathbb{R}^n,$$

with respect to the Lebesgue measure in $\mathbb{R}^n$. Hence the Radon-Nikodym derivative of μ_{E+h} with respect to μ_E is

$$\frac{\mathrm{d}\mu_{E+h}}{\mathrm{d}\mu_E}(y) = \exp\left(-\frac{h^T J^{-1} h - 2y^T J^{-1} h}{2\varepsilon^2}\right), \quad y \in \mathbb{R}^n,$$

for all $h \in \mathbb{R}^n$. Therefore also the normal distribution in $\mathbb{R}^n$ fulfills the assumptions of Theorem 2.2 for the noise distribution.

Since $R/\sim$ is homeomorphic to $\mathbb{R}^n$, by the definition it is a Polish space. Thus by Theorem 2.2 the posterior distribution μ_{post} of the roots $[Z]$ given the noisy real coefficients $a = (a_1, a_2, \dots, a_n)$ is

$$\mu_{\text{post}}(U) = \frac{\int_U \exp\left(-\frac{(F([z])-a)^T J^{-1}(F([z])-a)}{2\varepsilon^2}\right) \mu_{\text{pr}}(\mathrm{d}[z])}{\int_{R/\sim} \exp\left(-\frac{(F([z])-a)^T J^{-1}(F([z])-a)}{2\varepsilon^2}\right) \mu_{\text{pr}}(\mathrm{d}[z])}$$

for almost all $a \in \mathbb{R}^n$ where $U \in \mathcal{B}(R/\sim)$ and μ_{pr} is a prior distribution of the roots in $R/\sim$.

Similarly as in the complex case, we use the prior distribution μ_{pr} which is a combination of lifted uniform distributions taking into account all possible multiplicities of n roots, i. e.,

$$\mu_{\text{pr}}(U) = \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{m=1}^{l} \sum_{\substack{j_1+j_2+\ldots+j_m=l \\ 1\le j_1\le j_2\le\ldots\le j_m\le l}} \sum_{k=1}^{n-2l} \sum_{\substack{i_1+i_2+\ldots+i_k=n-2l \\ 1\le i_1\le i_2\le\ldots\le i_k\le n-2l}} \frac{\alpha_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_m}}{2^k \pi^m M^{k+2m}}$$
$$\times \int\limits_{\{(x_1,\ldots,x_k)\in\mathbb{R}^k,\ (w_1,\ldots,w_m)\in\mathbb{C}^m : [(x_{i_1,i_2,\ldots,i_k}, \bar{w}_{j_1,j_2,\ldots,j_m})]\in U\}} \chi_{[-M,M]^k}(x_1,\ldots,x_k)\chi_{B_0(M)^m}(w_1,\ldots,w_m)\,\mathrm{d}(x_1,\ldots,x_k,w_1,\ldots,w_m)$$

where $U \in \mathcal{B}(R/\sim)$, $B_0(M)$ is the disc in $\mathbb{C}$ with centre 0 and radius $M > 0$,

$$x_{i_1,i_2,\ldots,i_k} := (\underbrace{x_1,\ldots,x_1}_{i_1}, \underbrace{x_2,\ldots,x_2}_{i_2}, \ldots, \underbrace{x_k,\ldots,x_k}_{i_k}) \in \mathbb{R}^{n-2l}$$

for all $x \in \mathbb{R}^k$,

$$\bar{w}_{j_1,j_2,\ldots,j_m} := (\underbrace{w_1,\ldots,w_1}_{j_1}, \underbrace{\bar{w}_1,\ldots,\bar{w}_1}_{j_1}, \ldots, \underbrace{w_m,\ldots,w_m}_{j_m}, \underbrace{\bar{w}_m,\ldots,\bar{w}_m}_{j_m}) \in \mathbb{C}^{2l}$$

for all $w \in \mathbb{C}^m$, and $\alpha_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_m}$ are constants on the interval $[0, 1]$ satisfying

$$\sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{m=1}^{l} \sum_{\substack{j_1+j_2+\ldots+j_m=l \\ 1\le j_1\le j_2\le\ldots\le j_m\le l}} \sum_{k=1}^{n-2l} \sum_{\substack{i_1+i_2+\ldots+i_k=n-2l \\ 1\le i_1\le i_2\le\ldots\le i_k\le n-2l}} \alpha_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_m} = 1.$$

The constant $\alpha_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_m}$ is the prior probability of the appearance of k separate real roots with multiplicities i_1, i_2, $\ldots$, and i_k and m separate complex conjugate pairs with multiplicities j_1, j_2, …, and j_m for all $k = 1, \ldots, n-2l$ and $m = 1, \ldots, l$ where $l \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor\}$.

Let us denote $\|y\|_J^2 = y^T J^{-1} y$ for all $y \in \mathbb{R}^n$. If we want to evaluate the posterior probability of a given measurable set with the above choice of the prior distribution, we need to calculate integrals of the form

$$\int\limits_{\{x \in [-M,M]^k,\, w \in B_0(M)^m : [(x_{i_1,i_2,\ldots,i_k}, \bar{w}_{j_1,j_2,\ldots,j_m})] \in U\}} \exp\left(-\frac{1}{2\varepsilon^2} \|F([(x_{i_1,i_2,\ldots,i_k}, \bar{w}_{j_1,j_2,\ldots,j_m})]) - a\|_J^2 \right) \mathrm{d}(x, w)$$

for all $(i_1, i_2, \ldots, i_k)$ and $(j_1, j_2, \ldots, j_m)$ such that $1 \le i_1 \le i_2 \le \ldots \le i_k \le n - 2l$, $i_1 + i_2 + \ldots + i_k = n - 2l$, $1 \le j_1 \le j_2 \le \ldots \le j_m \le l$, and $j_1 + j_2 + \ldots + j_m = l$ for some $k = 1, \ldots, n-2l$ and $m = 1, \ldots, l$ where $l = 0, \ldots, \lfloor \frac{n}{2} \rfloor$ and $U \in \mathcal{B}(R/\sim)$.

For calculating posterior probabilities of different root configurations, let $(i_1, i_2, \ldots, i_k)$ and $(j_1, j_2, \ldots, j_m)$ be such that $1 \le i_1 \le i_2 \le \ldots \le i_k \le n-2l$, $i_1+i_2+\ldots+i_k = n-2l$, $1 \le j_1 \le j_2 \le \ldots \le j_m \le l$, and $j_1+j_2+\ldots+j_m = l$ for some $k = 1, \ldots, n-2l$ and $m = 1, \ldots, l$ where $l = 0, \ldots, \lfloor \frac{n}{2} \rfloor$. The posterior probability of the multiplicity pattern $(i_1, i_2, \ldots, i_k)$ and $(j_1, j_2, \ldots, j_m)$, i.e., for having k separate real roots with multiplicities i_1, i_2, …, and i_k and m complex conjugate pairs with multiplicities j_1, j_2, …, and j_m, is the posterior probability of the set $V_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_m} := \{[z] \in R/\sim : (x_{i_1,i_2,\ldots,i_k}, \bar{w}_{j_1,j_2,\ldots,j_m}) \in [z]$ for some $x \in \mathbb{R}^k$ and $w \in \mathbb{C}^m$ such that $x_p \ne x_q$ and $w_p \ne w_q$ for all $p \ne q\}$. Let us denote

$$p_{i_1,\ldots,i_k}^{j_1,\ldots,j_m} = \frac{\alpha_{i_1,\ldots,i_k}^{j_1,\ldots,j_m}}{2^k \pi^m M^{k+2m}} \int\limits_{[-M,M]^k \times B_0(M)^m} e^{-\frac{1}{2\varepsilon^2}\|F([(x_{i_1,i_2,\ldots,i_k}, \bar{w}_{j_1,j_2,\ldots,j_m})]) - a\|_J^2} \,\mathrm{d}(x, w). \tag{5}$$

Then

$$\mu_{\text{post}}(V_{i_1,i_2,\ldots,i_k}^{j_1,j_2,\ldots,j_m}) = \frac{p_{i_1,\ldots,i_k}^{j_1,\ldots,j_m}}{\sum_{t=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{u=1}^{l} \sum_{\substack{s_1+\ldots+s_u=t \\ 1 \le s_1 \le \ldots \le s_u \le t}} \sum_{v=1}^{n-2t} \sum_{\substack{r_1+\ldots+r_v=n-2t \\ 1 \le r_1 \le \ldots \le r_v \le n-2t}} p_{r_1,\ldots,r_v}^{s_1,\ldots,s_u}}.$$

Remark 3.2. Like in the complex case, the presented method does not depend on the choices of the prior and the noise distributions. The only requirement is that the noise distribution satisfies the assumptions of Theorem 2.2.

4 Numerical Examples

In this section we want to apply the presented Bayesian method for computing posterior probabilities of different multiplicity patterns for various real univariate polynomials. The first example is the cubic polynomial $P(z) = z^3 + a_1 z^2 + a_2 z + a_3$ where $a_1, a_2, a_3 \in \mathbb{R}$. The possible multiplicity patterns are three real single roots, a real double root and a real single root, a real triple root, and a complex conjugate pair and a real single root. With the notation of the previous section, we are interested in the posterior probabilities of the sets $V_{1,1,1}$, $V_{1,2}$, V_3, and V_1^1. We need to evaluate the corresponding intergals in (5), i. e.,

$$
\begin{aligned}
p_{1,1,1} &= \frac{\alpha_{1,1,1}}{8M^3} \int\limits_{[-M,M]^3} e^{-\frac{N^2(x_1+x_2+x_3+a_1)^2+N(x_1x_2+x_1x_3+x_2x_3-a_2)^2+(x_1x_2x_3+a_3)^2}{2\varepsilon^2 N^3}} \,\mathrm{d}(x_1,x_2,x_3),\\
p_{1,2} &= \frac{\alpha_{1,2}}{4M^2} \int\limits_{[-M,M]^2} e^{-\frac{N^2(2x_1+x_2+a_1)^2+N(x_1^2+2x_1x_2-a_2)^2+(x_1^2x_2+a_3)^2}{2\varepsilon^2 N^3}} \,\mathrm{d}(x_1,x_2),\\
p_3 &= \frac{\alpha_3}{2M} \int\limits_{[-M,M]} e^{-\frac{N^2(3x+a_1)^2+N(3x^2-a_2)^2+(x^3+a_3)^2}{2\varepsilon^2 N^3}} \,\mathrm{d}(x),\\
p_1^1 &= \frac{\alpha_1^1}{2\pi M^3} \int\limits_{[-M,M]\times B_0(M)} e^{-\frac{N^2(x+2\mathcal{R}ez+a_1)^2+N(2x\mathcal{R}ez+|z|^2-a_2)^2+(x|z|^2+a_3)^2}{2\varepsilon^2 N^3}} \,\mathrm{d}(x,z).
\end{aligned}
$$

Then the posterior probability of $V_{1,1,1}$ is

$$\mu_{\text{post}}(V_{1,1,1}) = \frac{p_{1,1,1}}{p_{1,1,1}+p_{1,2}+p_3+p_1^1}.$$

The posterior probabilities of $V_{1,2}$, V_3, and V_1^1 are calculated analogously.

We evaluated the above integrals by the default algorithm provided by Maple. For multidimensional integrals, this is [2], a deterministic algorithm with a globally adaptive subdivision scheme. In the cases where the result of integration is interesting, the intregand has just a quite sharp peak, namely at the nearest points to the pejorative manifold corresponding to the multiplicity pattern considered. In this example, there are just two pejorative manifolds, namely the nonsingular points of the disriminant surface

$$-27a_3^2 + 18a_1a_2a_3 + a_1^2a_2^2 - 4a_1^3a_3 - 4a_2^3 = 0$$

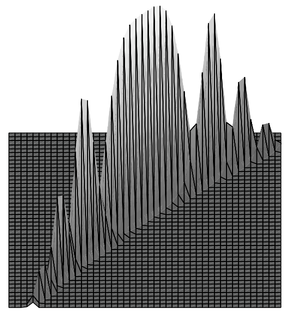

Figure 1: The plot of the function $(x, y) \mapsto e^{-100(x+y)^2-100(xy+5x+5y)^2}$ made by Maple to illustrate difficulties that Maple has with functions whose graph is a thin long ridge.

and the cubic space curve

$$a_1 = 3t,\ a_2 = 3t^2,\ a_3 = t^3,$$

along which the discriminant is singular. It is essential to translate that peak of the integrand to the origin, otherwise none of the cubature points lies on the peak and the result of the integration algorithm is meaningless.

In the cases where the integrand corresponds to the nearby pejorative manifold of highest codimension, the integrand is similar to a Gaussian density function with a covariance matrix which depends on the Jacobian of the function F. In borderline cases, this Jacobian becomes ill-conditioned, and the peak becomes thin and long in the directions of the singular vectors. When these vectors are not aligned with the coordinate axes, there are not enough points along the "ridge" of a long thin peak and hence the integration method does not reach convergence. A similar phenomenon can be observed when plotting functions. Instead of a ridge one sees a sequence of peaks, see Figure 1. Fortunately, this can be easily fixed: the coordinate system is rotated so that the axes are aligned with the singular vectors of the Jacobian matrix.

Another difficulty arises because these peaks are quite small compared to the integration box which sometimes leads to numerical instabilities in the integration algorithm. This can be improved by making the box smaller: in the directions where the function is similar to a Gaussian density function, we set the bounds equal to 5 times the expected error, and in the other directions we set the bounds so that the value of the integrand at the boundary is of the

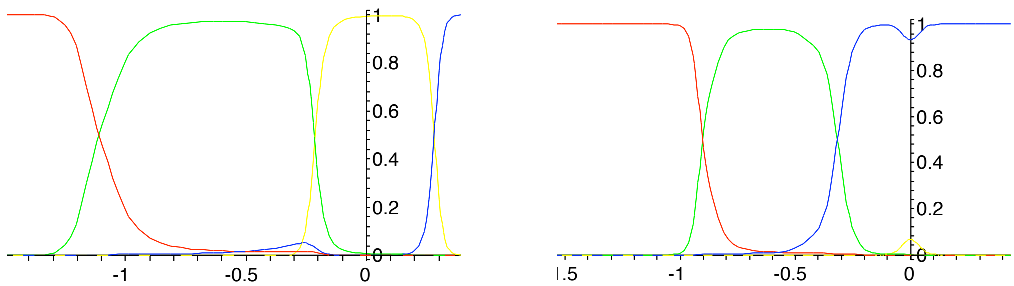

Figure 2: Posterior probabilities of different multiplicity patterns for the polynomial $P(z) = z^3 + az + 0.2$ when $\varepsilon = 0.1$ (left) and $\varepsilon = 0.05$ (right). Red denotes the posterior probability of three real single roots, green a double and a single root, yellow a triple root, and blue a single root and a complex conjugate pair.

order e^{-20}. In this way the computation becomes faster and more accurate on the cost of a neglectable error.

In Figure 2, the posterior probabilities of different multiplicity patterns for the polynomial $P(z) = z^3 + az + 0.2$ are presented as a function of the coefficient a. The prior probabilities $\alpha_{1,1,1}$, $\alpha_{1,2}$, α_3, and α_1^1 have all been set to 1/4. The other parameters are $\varepsilon = 0.1/0.05$, $N = 1$, and $M = 25$. A interesting conclusion is that for both values of the noise parameter ε, we can observe phase transitions: for most of the values of the coefficient a, there is only one multiplicity pattern with a high probability. Hence the presented method can be use to select a probable multiplicity pattern among all possible ones.

The same phenomenon is illustrated in Figure 3 for the polynomial $P(z) = z^3 - az + b$ when $(a, b) \in [-1, 4] \times [-4, 2]$. The selected parameters are $\alpha_{1,1,1} = \alpha_{1,2} = \alpha_3 = \alpha_1^1 = 1/4$, $\varepsilon = 0.5/0.25/0.125$, $N = 1$, and $M = 25$. We have chosen ε to be unrealistically large because of illustrative purposes. In order to tell which multiplicity pattern is the most likely one, it is useful to view the graphs of the posterior probability function from above. Seen from that perspective, the pejorative manifolds of a double and a single roots and a triple root seem to have a neighborhood of high probability of the corresponding multiplicity patter whose radius depends on the variance ε of the noise distribution.

In order to study more closely the effect of the parameters ε and M on the posterior probabilities of multiple roots, we compute the posterior probability

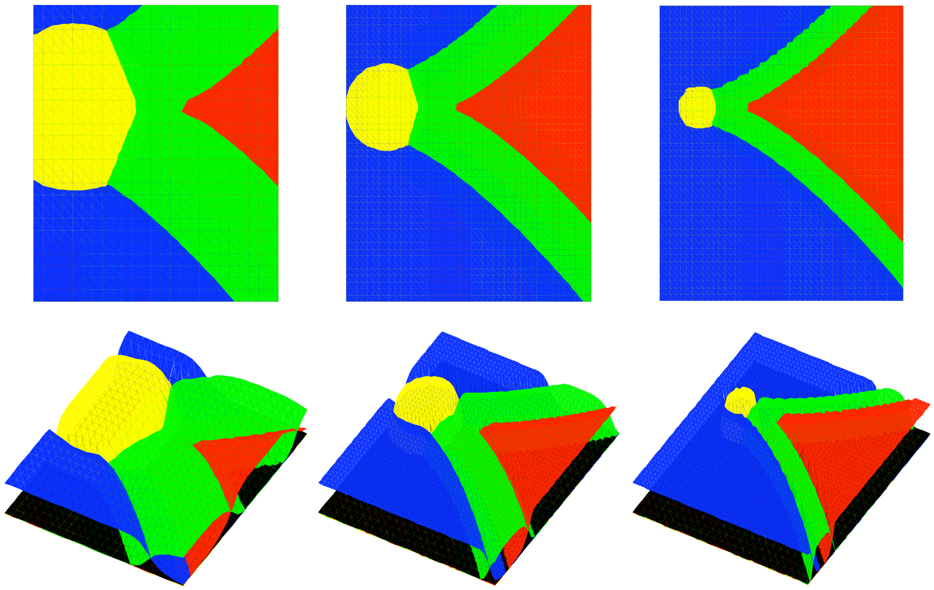

Figure 3: Posterior probabilities of different multiplicity patterns for the polynomial $P(z) = z^3 - az + b$ on the rectangle $(a, b) \in [-1, 4] \times [-4, 2]$ when $\varepsilon = 0.5$ (left), $\varepsilon = 0.25$ (middle) and $\varepsilon = 0.125$ (right). Red denotes the posterior probability of three real single roots, green a double and a single root, yellow a triple root, and blue a single root and a complex conjugate pair.

of a double root for a real polynomial $P(z) = z^2 + a_1 z + a_2$, i. e.,

$$\mu_{\text{post}}(V_2) = \frac{p_2}{p_{1,1} + p_2 + p^1}$$

where

$$p_{1,1} = \frac{\alpha_{1,1}}{4M^2} \int_{[-M,M]^2} e^{-\frac{N(x_1+x_2+a_1)^2 + (x_1 x_2 - a_2)^2}{2\varepsilon^2 N^2}} \, \mathrm{d}(x_1, x_2),$$

$$p_2 = \frac{\alpha_2}{2M} \int_{[-M,M]} e^{-\frac{N(2x+a_1)^2 + (x^2 - a_2)^2}{2\varepsilon^2 N^2}} \, \mathrm{d}x,$$

$$p^1 = \frac{\alpha^1}{\pi M^2} \int_{B_0(M)} e^{-\frac{N(\Re z + a_1)^2 + (|z|^2 - a_2)^2}{2\varepsilon^2 N^2}} \, \mathrm{d}z.$$

In the numerical examples we examine the polynomial $P(z) = z^2 + az + 4$. In the exact sense, P has a double root if and only if $a = 4$. In Figure 4, we

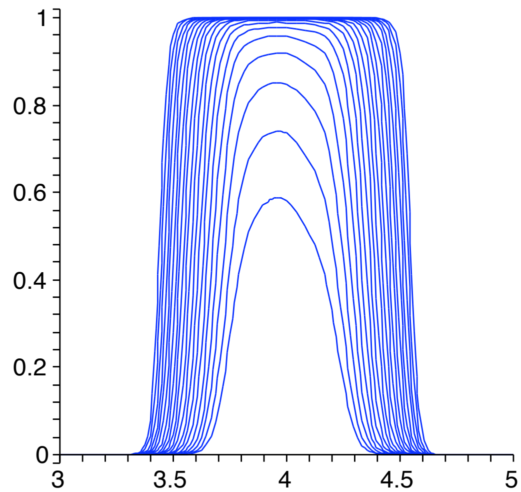

Figure 4: Posterior probability of a double root for the polynomial $P(z) = z^2 + az + 4$ when the prior bound M is $M = 2^k$, $k = 1, \dots, 20$.

set $\varepsilon = 2^{-3}$ and $M = 2^k$ for $k = 1, \dots, 20$. The other parameters are $\alpha_{1,1} = \alpha_2 = \alpha^1 = 1/3$, and $N = 1$. As M increases, the posterior probability of a double root also increases. However, the graphs of the posterior probability functions for the parameters $M = 2^8, 2^9, \dots, 2^{20}$ look qualitatively similar. In Figure 5, we set $M = 2^4$ and $\varepsilon = 2^{-k}$ for $k = 1, \dots, 6$. As ε decreases, the region where $\mu_{\text{post}}(V_2)$ is large becomes smaller but at the same time the maximum of the posterior probability becomes larger. The changes are more dramatic than the changes for M. Figure 5 also shows that the maximum of $\mu_{\text{post}}(V_2)$ is not achieved at $a = 4$ but at some slightly smaller value depending on ε. At the moment we do not have an explanation for this phenomenon.

As a last example we study an example used in [12, Section 5.2]. The real polynomial

$$P_a(z) = (z - 1 + a)^{20}(z - 1)^{20}(z + 0.5)^5, \; a = 10^{-l}, \; l = 1, 2, 3, \dots,$$

has two nearby roots of multiplicity 20. The number of possible multiplicity patterns is 1 353 106 but one can see that if the prior probabilities $\alpha^{j_1, j_2, \dots, j_m}_{i_1, i_2, \dots, i_k}$ are all roughly equal, the posterior probabilities are dominated by two cases, namely $(5, 20, 20)$ and $(5, 40)$. Therefore we need to evaluate only the inte-

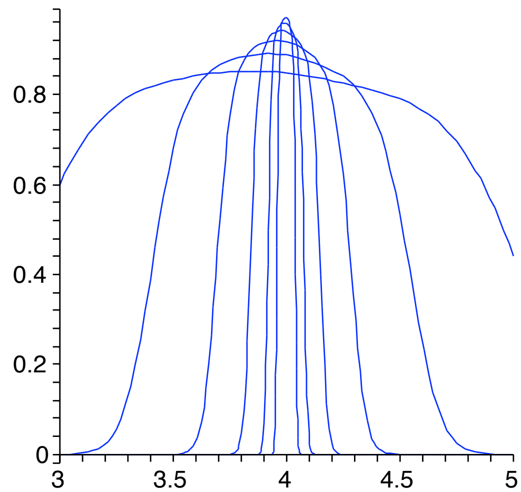

Figure 5: Posterior probability of a double root for the polynomial $P(z) = z^2 + az + 4$ when the variance ε of the noise distribution is $\varepsilon = 2^{-k}$, $k = 1, \ldots, 6$.

grals

$$p_{5,20,20} = \frac{\alpha_{5,20,20}}{8M^3} \int_{[-M,M]^3} e^{-\frac{||F([(x_1,x_2,x_3)_{5,20,20}]) - F([(0.5,1,1+a)_{5,20,20}])||_J^2}{2\varepsilon^2}} d(x_1, x_2, x_3)$$

and

$$p_{5,40} = \frac{\alpha_{5,40}}{4M^2} \int_{[-M,M]^2} e^{-\frac{||F([(x_1,x_2)_{5,40}]) - F([(0.5,1,1+a)_{5,20,20}])||_J^2}{2\varepsilon^2}} d(x_1, x_2).$$

The posterior probabilities can be approximated by

$$\mu_{\text{post}}(V_{5,20,20}) \approx \mu_1 := \frac{p_{5,20,20}}{p_{5,20,20} + p_{5,40}}, \quad \mu_{\text{post}}(V_{5,40}) \approx \mu_2 := \frac{p_{5,40}}{p_{5,20,20} + p_{5,40}}.$$

We fix the parameter N to be $N = 5$ in order to compensate that the higher coefficients are quite large, and the a priori bound to be $M = 20$. The posterior probabilities for various values of a are given in Table 1. A clear observation is that the posterior probability μ_2 increases as a decreases for all values of ε. The results in [12] show for small a the use of the multiplicity

Table 1: The posterior probabilities of the multiplicity patterns $(5, 20, 20)$ and $(5, 40)$ for the polynomial $P_a(z) = (z - 1 + a)^{20}(z - 1)^{20}(z + 0.5)^5$ for different values of a and ε. Only the smaller value of $\mu_1 \approx \mu_{\text{post}}(V_{5,20,20})$ and $\mu_2 \approx \mu_{\text{post}}(V_{5,40})$ is written since $\mu_1 = 1 - \mu_2$.

	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-7}$	$\varepsilon = 10^{-8}$
$a = 10^{-3}$	$\mu_2 \approx 10^{-226}$	$\mu_2 < 10^{-999}$	$\mu_2 < 10^{-999}$	$\mu_2 < 10^{-999}$
$a = 10^{-4}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 0.068$	$\mu_2 \approx 10^{-225}$	$\mu_2 < 10^{-999}$
$a = 10^{-5}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-4}$	$\mu_1 \approx 0.0072$
$a = 10^{-6}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-4}$	$\mu_1 \approx 10^{-4}$
$a = 10^{-7}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-3}$	$\mu_1 \approx 10^{-4}$	$\mu_1 \approx 10^{-4}$

pattern $(5, 20, 20)$ leads to a larger backward error than the multiplicity pattern $(5, 40)$. Hence the more probable multiplicity pattern is also the pattern of the greater significance from a numeric point of view.

5 Conclusions

In this paper, we have studied the root computation problem in the inverse problem context. The main purpose of the paper is to introduce a statistical method to estimate the root configuration of an univariate polynomial based on numerically given coefficients. A Bayesian method for calculating the posterior probabilities of different multiplicity patters has been presented. A specific choice of the noise and the prior distributions has been utilized. The method itself does not depend on these choices but allows application oriented noise and prior distributions to be used in the computation of the posterior probabilities of root configurations. The method has been illustrated by several numerical examples. The examples show that the presented method can be used to select a probable multiplicity pattern among all possible ones and is in the consistence with previous results.

Acknowledgements

The second author was partially supported by the FWF (Austrian science fund) in the frame of the doctoral programm "Computational Mathematics". We also thank A. Neubauer for helpful remarks on numerical integration.

References

[1] H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen. *Linear and Graphical Models.* Springer-Verlag, New York, U.S.A., 1995.

[2] J. Berntsen, T. Espelid, and A. Genz. An adaptive algorithm for the approximate calculation of multiple integrals. *ACM TOMS*, 17:437–451, 1991.

[3] D. Bini and G. Fiorentino. Design, analysis, and implementation of a multiprecision polynomial rootfinder. *Num. Alg.*, 23:127–173, 2000.

[4] R. Corless, P. Chin, and G. Corliss. Optimization strategies for the approximate GCD problem. In *Proc. ISSAC'98*, pages 228–235. ACM Press, 1998.

[5] R. Corless, P. Gianni, B. Trager, and S. Watt. The singular value decomposition for polynomial systems. In *Proc. ISSAC'95*, pages 195–207. ACM Press, 1995.

[6] R. M. Dudley. *Real Analysis and Probability.* Cambridge University Press, New York, U.S.A., 2002.

[7] V. Hribernig and H.-J. Stetter. Detection and validation of clusters of polynomial zeroes. *J. Symb. Comp.*, 24:667–682, 1997.

[8] J. P. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems.* Springer-Verlag, Berlin, Germany, 2004.

[9] M. Lassas, E. Saksman, and S. Siltanen. Discretization invariant Bayesian inversion and Besov space prior. *submitted.*

[10] V. Y. Pan. Univariate polynomials: nearly optimal algorithms for numerical factorization and root-finding. *J. Symb. Comp.*, 33:701–733, 2002.

[11] J.-C. Yakoubsohn. Finding a cluster of zeroes of univariate polynomials. *J. Complexity*, 16:603–638, 2000.

[12] Z. Zeng. Computing multiple roots of inexact polynomials. *Math. Comp.*, 74:869–903, 2005.