



Article

Regression with Gaussian Mixture Models Applied to Track Fitting

Rudolf Frühwirth

Institute of High Energy Physics (HEPHY), Austrian Academy of Sciences, Nikolsdorfer Gasse 18, 1050 Wien, Austria; rudolf.fruehwirth@oeaw.ac.at

Received: 28 July 2020; Accepted: 28 August 2020; Published: 31 August 2020

Abstract: This note describes the application of Gaussian mixture regression to track fitting with a Gaussian mixture model of the position errors. The mixture model is assumed to have two components with identical component means. Under the premise that the association of each measurement to a specific mixture component is known, the Gaussian mixture regression is shown to have consistently better resolution than weighted linear regression with equivalent homoskedastic errors. The improvement that can be achieved is systematically investigated over a wide range of mixture distributions. The results confirm that with constant homoskedastic variance the gain is larger for larger mixture weight of the narrow component and for smaller ratio of the width of the narrow component and the width of the wide component.

Keywords: track fit; Gaussian mixture model; linear regression

1. Introduction

In the data analysis of high-energy physics experiments, and particularly in track fitting, Gaussian models of measurements errors or Gaussian models of stochastic processes such as multiple Coulomb scattering and energy loss of electrons by bremsstrahlung may turn out to be too simplistic or downright invalid. In these circumstances Gaussian mixture models (GMMs) can be used in the data analysis to model outliers or tails in position measurements of tracks [1,2], to model the tails of the multiple Coulomb scattering distribution [3,4], or to approximate the distribution of the energy loss of electrons by bremsstrahlung [5]. In these applications it is not known to which component an observation or a physical process corresponds, and it is up to the track fit via the Deterministic Annealing Filter or the Gaussian-sum Filter to determine the association probabilities [6–9]. The standard method for concurrent maximum-likelihood estimation of the regression parameters and the association probabilities is the EM algorithm [10,11].

In a typical tracking detector such as a silicon strip tracker, the spatial resolution of the measured position is not necessarily uniform across the sensors and depends on the crossing angle of the track and on the type of the cluster created by the particle. By a careful calibration of the sensors in the tracking detector a GMM of the position measurements can be formulated, see [12–14]. In such a model the component from which an observation is drawn is known. This can significantly improve the precision of the estimated regression/track parameters with respect to weighted least-squares, as shown below.

This introduction is followed by a recapitulation of weighted regression in linear and nonlinear Gaussian models in Section 2, both for the homoskedastic and for the heteroskedastic case. Section 3 introduces the concept of Gaussian mixture regression (GMR) and derives some elementary results on the covariance matrices of the estimated regression parameters. Section 4 presents the results of simulation studies that apply GMR to track fitting in a simplified tracking detector. The gain in precision that can be obtained by using a GMM rather than a simple Gaussian model strongly depends

on the shape of the mixture distribution of the position errors. This shape can be described concisely by two characteristic quantities: the weight p of the wider component, and the ratio r of the smaller over the larger standard deviation. The study covers a wide range of p and r and shows the gain as a function of the number of measurements within a fixed tracker volume. It also investigates the effect of multiple Coulomb scattering on the gain in precision. Finally, Section 5 gives a summary of the results and discusses them in context of the related work in [12–14].

2. Linear Regression in Gaussian Models

This section briefly recapitulates the properties of linear Gaussian regression models and the associated estimation procedures.

2.1. Linear Gaussian Regression Models

The general linear Gaussian regression model (LGRM) has the following form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{c} + \boldsymbol{\epsilon}, \quad \text{with } \mathbf{V} = \text{Var}[\boldsymbol{\epsilon}] \quad \text{and} \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}), \quad (1)$$

where \mathbf{y} is the vector of observations, $\boldsymbol{\theta}$ is the vector of unknown regression parameters, \mathbf{X} is the known model matrix, \mathbf{c} is a known vector of constants, and $\boldsymbol{\epsilon}$ is the vector of observation errors, distributed according to the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{V})$ with mean $\mathbf{0}$ and covariance matrix \mathbf{V} . In the following it is assumed that $\dim(\mathbf{y}) \geq \dim(\boldsymbol{\theta})$ and $\text{rank}(\mathbf{X}) = \dim(\boldsymbol{\theta})$.

If the covariance matrix \mathbf{V} is diagonal, i.e., the observation errors are independent, two cases are distinguished in the literature. The errors in Equation (1) and the model are called *homoskedastic*, if \mathbf{V} is a multiple of the identity matrix:

$$\mathbf{V} = \sigma^2 \cdot \mathbf{I}. \quad (2)$$

Otherwise, the errors in Equation (1) and the model are called *heteroskedastic*. If \mathbf{V} is not diagonal, the observation errors are correlated, in which case the model is called homoskedastic if all diagonal elements are identical, and heteroskedastic otherwise.

2.2. Estimation, Fisher Information Matrix and Efficiency

In least-squares (LS) estimation, the regression parameters $\boldsymbol{\theta}$ are estimated by minimizing the following quadratic form with respect to $\boldsymbol{\theta}$:

$$\mathcal{S}(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta})^T \mathbf{G}(\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta}), \quad \text{with } \mathbf{G} = \mathbf{V}^{-1}. \quad (3)$$

Setting the gradient of $\mathcal{S}(\boldsymbol{\theta})$ to zero gives the LS estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and linear error propagation yields its covariance matrix $\hat{\mathbf{C}}$:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{G} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\mathbf{y} - \mathbf{c}), \quad \hat{\mathbf{C}} = \text{Var}[\hat{\boldsymbol{\theta}}] = (\mathbf{X}^T \mathbf{G} \mathbf{X})^{-1}. \quad (4)$$

It is easy to see that $\hat{\boldsymbol{\theta}}$ is an unbiased estimate of the unknown parameter $\boldsymbol{\theta}$.

In the LGRM, the LS estimate is equal to the maximum-likelihood (ML) estimate. The likelihood function $\mathcal{L}(\boldsymbol{\theta})$ and its logarithm $\ell(\boldsymbol{\theta})$ are given by:

$$\mathcal{L}(\boldsymbol{\theta}) = C \cdot \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta})^T \mathbf{G}(\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta}) \right], \quad (5)$$

$$\ell(\boldsymbol{\theta}) = \log(C) - \frac{1}{2} (\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta})^T \mathbf{G}(\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta}), \quad C \in \mathbb{R}. \quad (6)$$

The gradient $\nabla\ell(\boldsymbol{\theta})$ is equal to:

$$\nabla\ell(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = (\mathbf{y} - \mathbf{c} - \mathbf{X}\boldsymbol{\theta})^T \mathbf{G}\mathbf{X}, \tag{7}$$

and its matrix of second derivatives, the Hesse matrix $\mathbf{H}(\boldsymbol{\theta})$ is equal to:

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2} = -\mathbf{X}^T \mathbf{G}\mathbf{X}. \tag{8}$$

Setting the gradient to zero yields the ML estimate, which is identical to the LS estimate.

Under mild regularity conditions, the covariance matrix of an unbiased estimator cannot be smaller than the inverse of the Fisher information matrix; this is the famous Cramér-Rao inequality [15]. The (expected) Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ of the parameters $\boldsymbol{\theta}$ is defined as the negative expectation of the Hesse matrix of the log-likelihood function:

$$\mathcal{I}(\boldsymbol{\theta}) = -E[\mathbf{H}(\boldsymbol{\theta})] = \mathbf{X}^T \mathbf{G}\mathbf{X}. \tag{9}$$

Equation (4) shows that in the LGRM the covariance matrix of the LS/ML estimate is *equal* to the inverse of the Fisher information matrix. The LS/ML estimator is therefore *efficient*: no other unbiased estimator of $\boldsymbol{\theta}$ can have a smaller covariance matrix.

2.3. Nonlinear Gaussian Regression Models

In most applications to track fitting, the track model that describes the dependence of the observed track positions on the track parameters is nonlinear. The general nonlinear regression Gaussian model has the following form:

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}), \tag{10}$$

where $\mathbf{f}(\boldsymbol{\theta})$ is a nonlinear smooth function. The estimation of $\boldsymbol{\theta}$ usually proceeds via Taylor expansion of the model function to first order:

$$\mathbf{y} \approx \mathbf{f}(\boldsymbol{\theta}_0) + \mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\theta} + \mathbf{c} + \boldsymbol{\epsilon}, \quad \text{with} \tag{11}$$

$$\mathbf{X} = \left. \frac{\partial\mathbf{f}}{\partial\boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_0}, \quad \mathbf{c} = \mathbf{f}(\boldsymbol{\theta}_0) - \mathbf{X}\boldsymbol{\theta}_0, \tag{12}$$

where $\boldsymbol{\theta}_0$ is a suitable expansion point and the Jacobian \mathbf{X} is evaluated at $\boldsymbol{\theta}_0$. The estimate $\boldsymbol{\theta}_1$ is obtained according to Equation (4). The model function is then re-expanded at $\boldsymbol{\theta}_1$, and the procedure is iterated until convergence.

3. Linear Regression in Gaussian Mixture Models

If the observation errors do not follow a normal distribution, the Gaussian linear model can be generalized to a GMM, which is more flexible, but still retains some of the computational advantages of the Gaussian model. Applications of Gaussian mixtures can be traced back to Karl Pearson more than a century ago [16]. In view of the applications in Section 4, the following discussion is restricted to Gaussian mixtures with two components with identical means. The probability density function (PDF) $g(\mathbf{y})$ of such a mixture has the following form:

$$g(\mathbf{y}) = p \cdot \varphi(\mathbf{y}; \mu, \sigma_1) + (1 - p) \cdot \varphi(\mathbf{y}; \mu, \sigma_0), \quad 0 < p < 1, \quad \sigma_0 < \sigma_1, \tag{13}$$

where $\varphi(\mathbf{y}; \mu, \sigma)$ denotes the normal PDF with mean μ and variance σ^2 . The Gaussian mixture (GM) with this PDF is denoted by $\mathcal{GM}(p, \mu, \sigma_1, \sigma_0)$. The component with index 0 is called the narrow component, the component with index 1 is called the wide component in the following.

3.1. Homoskedastic Mixture Models

We first consider the following linear GMM:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{c} + \boldsymbol{\epsilon}, \quad \text{with } \epsilon_i \sim \mathcal{GM}(p, 0, \sigma_1, \sigma_0), \quad i = 1, \dots, n, \quad (14)$$

where $n = \dim(\mathbf{y})$. The variance of ϵ_i is equal to:

$$\text{var}[\epsilon_i] = \sigma_{\text{tot}}^2 = p \cdot \sigma_1^2 + (1 - p) \cdot \sigma_0^2, \quad i = 1, \dots, n, \quad (15)$$

and the joint covariance matrix of $\boldsymbol{\epsilon}$ is equal to

$$\mathbf{V} = \text{Var}[\boldsymbol{\epsilon}] = \sigma_{\text{tot}}^2 \cdot \mathbf{I}. \quad (16)$$

As the total variance is the same for all i , the GMM is called homoskedastic. By using just the total variances, it can be approximated by a homoskedastic LGRM, and the estimation of the regression parameters proceeds as in Equation (4), which can be simplified to:

$$\check{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{y} - \mathbf{c}), \quad \check{\mathbf{C}} = \sigma_{\text{tot}}^2 \cdot (\mathbf{X}^T \mathbf{X})^{-1}. \quad (17)$$

If the association of the observations to the components is known, it can be encoded in a binary vector $\mathbf{z} \in Z = \{0, 1\}^n$, where $z_i = j$ indicates the component with variance σ_j^2 , for $i = 1, \dots, n$. In the case of independent draws from the mixture PDF in Equation (13), the z_i are independent Bernoulli variables with $P(1) = p, P(0) = 1 - p$. The probability of $\mathbf{z} = (z_1, \dots, z_n)$ is therefore equal to:

$$P(\mathbf{z}) = \prod_{i=1}^n p^{z_i} (1 - p)^{1 - z_i} = p^k (1 - p)^{n - k}, \quad \text{with } k = \sum_{i=1}^n z_i. \quad (18)$$

Conditional on the known associations \mathbf{z} , the model can be interpreted as a heteroskedastic LGRM with a diagonal covariance matrix \mathbf{V}_z :

$$\mathbf{V}_z = \text{Var}[\boldsymbol{\epsilon} | \mathbf{z}] = \text{diag}(\sigma_{z_1}^2, \dots, \sigma_{z_n}^2), \quad \text{and } \mathbf{G}_z = \mathbf{V}_z^{-1}. \quad (19)$$

It follows from the binomial theorem that:

$$\mathbf{V} = \sum_{\mathbf{z} \in Z} P(\mathbf{z}) \cdot \mathbf{V}_z. \quad (20)$$

The GMR is thus a mixture of heteroskedastic Gaussian linear regressions with mixture weights $P(\mathbf{z}), \mathbf{z} \in Z$. In each of the heteroskedastic LGRMs the regression parameters are estimated by:

$$\hat{\boldsymbol{\theta}}_z = (\mathbf{X}^T \mathbf{G}_z \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}_z (\mathbf{y} - \mathbf{c}), \quad \hat{\mathbf{C}}_z = (\mathbf{X}^T \mathbf{G}_z \mathbf{X})^{-1}. \quad (21)$$

Conditional on \mathbf{z} , the estimator is normally distributed with mean $\boldsymbol{\theta}$ and covariance matrix $\hat{\mathbf{C}}_z$, and therefore unbiased and efficient. The unconditional distribution of $\hat{\boldsymbol{\theta}}$ is obtained by summing over all \mathbf{z} with weights $P(\mathbf{z})$. Its density therefore reads:

$$f(\hat{\boldsymbol{\theta}}) = \sum_{\mathbf{z} \in Z} P(\mathbf{z}) \cdot \varphi(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}, \hat{\mathbf{C}}_z), \quad (22)$$

where $\varphi(\hat{\theta}; \theta, \hat{C}_z)$ is the multivariate normal density of $\hat{\theta}$ with mean θ and covariance matrix \hat{C}_z . $\hat{\theta}$ is therefore unbiased, and its covariance matrix is obtained by summing over the 2^n possible values of z with weights $P(z)$:

$$\hat{C} = \sum_{z \in Z} P(z) \cdot \hat{C}_z. \tag{23}$$

The unconditional estimator has minimum variance among all estimators that are conditionally unbiased for all $z \in Z$, as shown by the following theorem.

Theorem 1. *Let $\tilde{\theta}$ be an estimator of θ with $E[\tilde{\theta}|z] = \theta$ for all $z \in Z$. Then its covariance matrix \tilde{C} is not smaller than \hat{C} in the Loewner partial order:*

$$\hat{C} \leq \tilde{C}. \tag{24}$$

Proof. As $E[\tilde{\theta}|z] = \theta$, \tilde{C} can be written in the following form:

$$\tilde{C} = \sum_{z \in Z} P(z) \cdot \tilde{C}_z, \quad \text{with} \quad \tilde{C}_z = \text{Var}[\tilde{\theta}|z]. \tag{25}$$

As $\hat{\theta}_z$ is efficient conditional on z and $\tilde{\theta}|z$ is conditionally unbiased, it follows that

$$\hat{C}_z = \text{Var}[\hat{\theta}|z] \leq \text{Var}[\tilde{\theta}|z] = \tilde{C}_z, \quad \text{for all } z \in Z. \tag{26}$$

Summing over all z with weights $P(z)$ proves the theorem. \square

Collorary. $\hat{C} \leq \check{C}$, with equality if $p \in \{0, 1\}$.

Proof. As $E[\check{\theta}|z] = \theta$ for all z , the estimator $\check{\theta}$ fulfills the premise of the theorem. Its conditional covariance matrix is obtained by linear error propagation:

$$\check{C}_z = \text{Var}[\check{\theta}|z] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_z \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \tag{27}$$

As $\hat{\theta}|z$ is efficient, $\hat{C}_z \leq \check{C}_z$ and $\hat{C} \leq \check{C}$. If $p \in \{0, 1\}$, \mathbf{V}_z is homoskedastic and $\hat{C}_z = \check{C}_z = \sigma_{\text{tot}}^2 \cdot (\mathbf{X}^T \mathbf{X})^{-1}$. \square

The effective improvement in precision of the heteroskedastic estimator $\hat{\theta}$ w.r.t. regarding to the homoskedastic estimator $\check{\theta}$ is difficult to quantify by simple formulas, but easy to assess by simulation studies. The results of three such studies in the context of track fitting are presented in Sections 4.1–4.3, respectively.

3.2. Heteroskedastic Mixture Models

The GMM in Equation (14) can be further generalized to the heteroskedastic case, where the mixture parameters depend on i :

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{c} + \boldsymbol{\epsilon}, \quad \text{with} \quad \epsilon_i \sim \mathcal{GM}(p_i, 0, \sigma_{i,1}, \sigma_{i,0}), \quad i = 1, \dots, n. \tag{28}$$

The joint covariance matrix of $\boldsymbol{\epsilon}$ is now diagonal with

$$\mathbf{V} = \text{Var}[\boldsymbol{\epsilon}] = \text{diag}(\sigma_{1,\text{tot}}^2, \dots, \sigma_{n,\text{tot}}^2), \quad \text{with} \tag{29}$$

$$\sigma_{i,\text{tot}}^2 = p_i \cdot \sigma_{i,1}^2 + (1 - p_i) \cdot \sigma_{i,0}^2, \quad i = 1, \dots, n. \tag{30}$$

The approximating LGRM that uses just the total variances is now heteroskedastic with covariance matrix \mathbf{V} . The corresponding estimator $\check{\theta}$ with its covariance matrix \check{C} is computed as in Equation (4).

Conditional on the associations coded in z , the covariance matrix V_z now reads:

$$V_z = \text{Var}[e|z] = \text{diag}(\sigma_{1,z_1}^2, \dots, \sigma_{n,z_n}^2). \tag{31}$$

As in Section 3.1, the unconditional covariance matrix \hat{C} is the weighted sum of the conditional covariance matrices \hat{C}_z :

$$\hat{C} = \sum_{z \in Z} P(z) \cdot \hat{C}_z, \quad \text{with} \quad P(z) = \prod_{i=1}^n p_i^{z_i} (1 - p_i)^{1-z_i}. \tag{32}$$

It can be proved as before that $\hat{C} \leq \check{C}$.

In the application to track fitting, heteroskedastic mixture models can be useful in non-uniform tracking detectors, in which the mixture model obtained by the calibration depends on the layer or on the sensor or on the possible cluster types.

A nonlinear GMM can be approximated by a linear GMM in the same way as a nonlinear Gaussian model can be approximated by a LGRM, see Section 2.3.

4. Track Fitting with Gaussian Mixture Models

4.1. Simulation Study with Straight Tracks

The setting of the first simulation study is a toy detector that consists of n equally spaced sensor planes in the interval from 0 cm to 115 cm. The track model is a straight line in the (x, y) -plane. No material effects are simulated. The track position is recorded in each sensor plane, with a position error generated from a Gaussian mixture, see Equation (13). The total variance in each layer is equal to $\sigma_{\text{tot}}^2 = (0.005 \text{ cm})^2$, corresponding to a standard deviation of $\sigma_{\text{tot}} = 0.005 \text{ cm}$. As the main objective of the simulation study is the comparison of the homoskedastic estimator of Equation (17) with the heteroskedastic estimator of Equation (21), the actual value of σ_{tot}^2 is irrelevant.

10^5 tracks were generated for each of 20 combinations of p and $r = \sigma_0/\sigma_1$: $p = 0.6, 0.7, 0.8, 0.9$; $r = 0.1, 0.2, 0.3, 0.4, 0.5$. The corresponding values of σ_1 and σ_0 are collected in Table 1. The number n of planes varied from 3 to 20. For each pair (p, r) and each value of n the exact covariance matrix was computed according to Equation (23).

Table 1. Values of σ_1 and σ_0 (in μm) for $p = 0.6, 0.7, 0.8, 0.9$ and $r = 0.1, 0.2, 0.3, 0.4, 0.5$.

	$p = 0.6$		$p = 0.7$		$p = 0.8$		$p = 0.9$	
	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0
$r = 0.1$	64.3	6.4	59.6	6.0	55.8	5.6	52.7	5.3
$r = 0.2$	63.7	12.7	59.3	11.9	55.6	11.1	52.6	10.5
$r = 0.3$	62.7	18.8	58.6	17.6	55.3	16.6	52.4	15.7
$r = 0.4$	61.4	24.5	57.8	23.1	54.8	21.9	52.2	20.9
$r = 0.5$	59.7	29.9	56.8	28.4	54.2	27.1	52.0	26.0

The track parameters were chosen as the intercept d and the slope k . All tracks were fitted with the homoskedastic and with the heteroskedastic error model. Figure 1 shows the estimated standard deviation (rms) of both fits, along with the exact values, for the intercept d and the slope k , with $p = 0.8$, $r = 0.2$ in the range $3 \leq n \leq 20$. The standard deviations of the heteroskedastic fit are not only smaller, but also fall at a faster rate than the ones of the homoskedastic fit, i.e., faster than $1/\sqrt{n}$. The agreement between the estimated and the exact standard deviations is excellent.

Figure 2 shows the relative standard deviation (heteroskedastic rms over homoskedastic rms) for the slope k and all 20 combinations of p and r . The corresponding plot for the intercept d looks very similar. A clear trend can be observed: the ratio gets smaller with decreasing p and with decreasing r .

The largest improvement is therefore seen for $p = 0.6, r = 0.1$, where the probability of very precise measurements ($r = 0.1$) is the largest ($1 - p = 0.4$).

For large values of n the ratio curves approach a constant level, because asymptotically both rms values fall at the same rate, namely $1/\sqrt{n}$, see Figure 3. In the experimentally relevant region $15 \leq n \leq 20$ the ratio is already close to its lower limit in most, but not all, cases. In these cases, the superior performance of the heteroskedastic fit is fully exploited, and adding more measurements yields little further improvement. The exception is the case $p = 0.9$, where the limit is reached much later, at $n = 200$ in the extreme case.

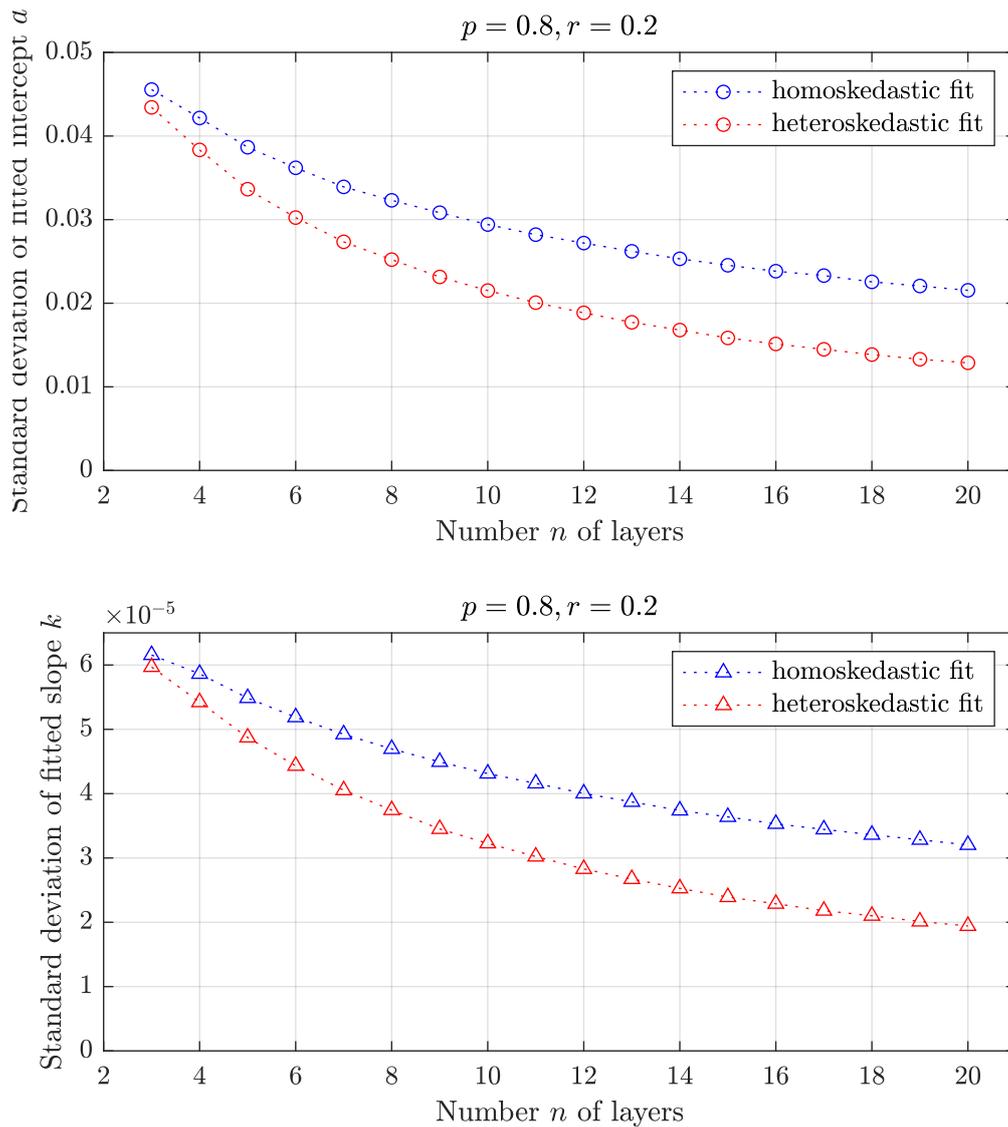


Figure 1. Top: estimated rms and exact standard deviation of the fitted intercept d ; bottom: estimated rms and exact standard deviation of the fitted slope k , for $p = 0.8, r = 0.2$ in the range $3 \leq n \leq 20$.

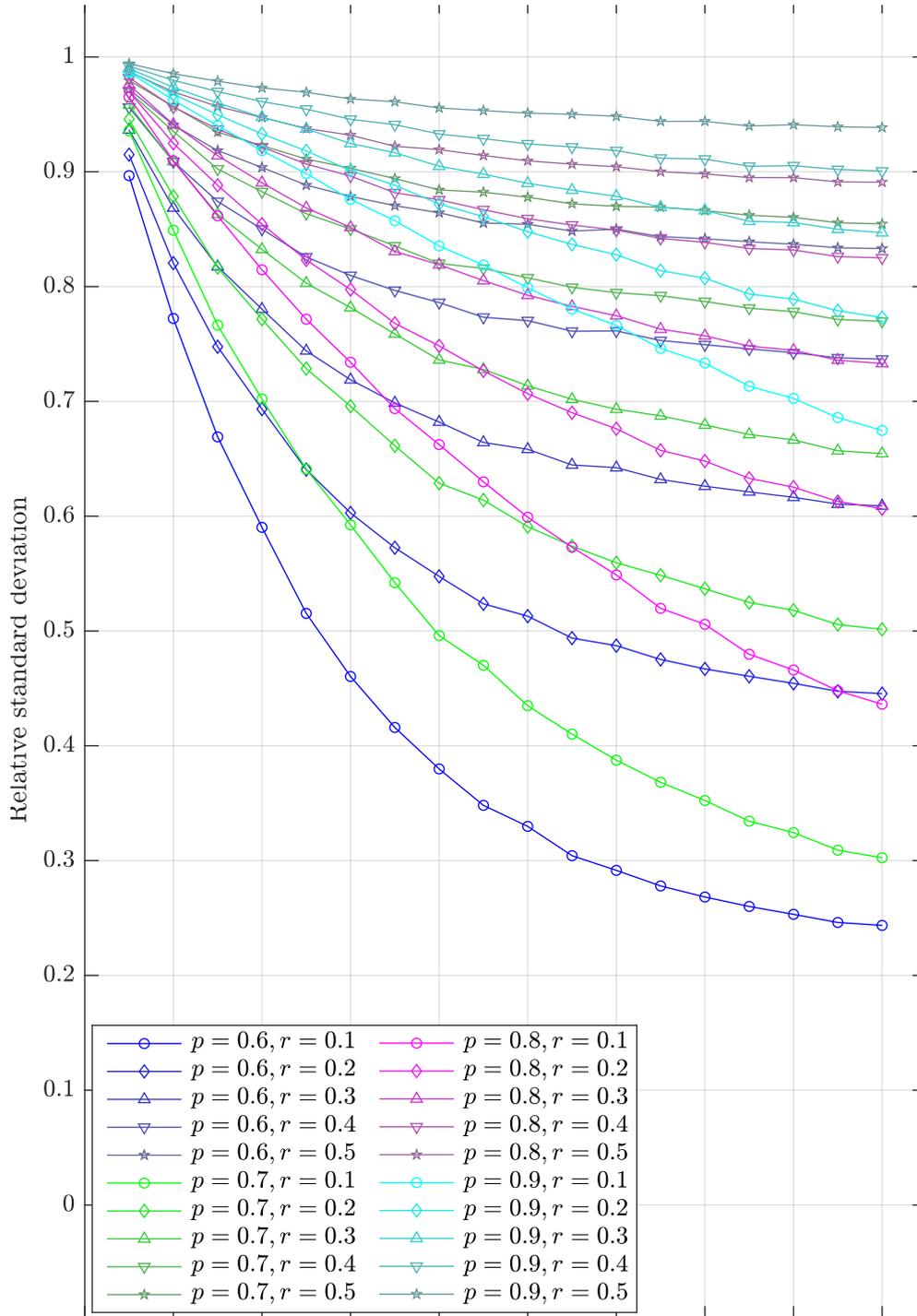


Figure 2. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the fitted slope k for all 20 combinations of p and r in the range $3 \leq n \leq 20$.

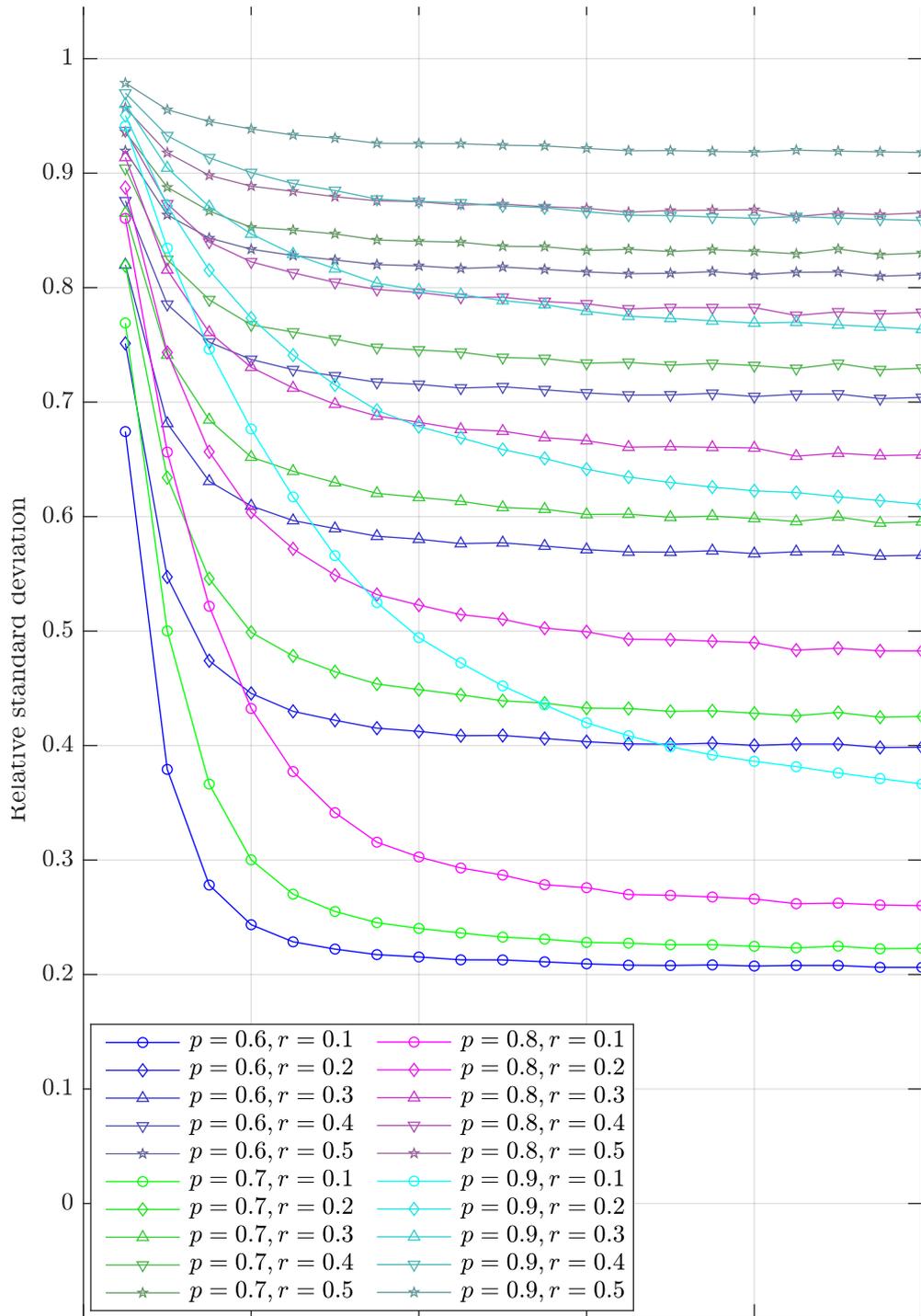


Figure 3. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the fitted slope k for all 20 combinations of p and r in the range $5 \leq n \leq 100$.

4.2. Simulation Study with Circular Tracks

The setting of the second simulation study is a toy detector that consists of n equally spaced concentric cylinders with radii in the interval $[R_{\min}, R_{\max}]$, with $R_{\min} = 10$, $R_{\max} = 100$ cm. The track model is a helix through the origin. Only the parameters in the bending plane ((x, y) -plane) are estimated. The track parameters are: Φ , the azimuth of the track position at $R = R_{\min}$; ϕ , the azimuth of the track direction at $R = R_{\min}$; and κ , the curvature of the circle.

10^6 tracks were generated for each of the same 20 combinations of p and $r = \sigma_0/\sigma_1$ as in Section 4.1. The track position in the bending plane was recorded in each cylinder, with a position error generated from the Gaussian mixtures in Table 1. Each sample of 10^6 tracks consists of 10 subsamples with circles of 10 different radii, namely $\rho/\text{cm} = 100, 200, \dots, 1000$. The number n of planes varied from 3 to 20.

As the track model is nonlinear, estimation was performed according to the procedure in Section 2.3. All tracks were fitted with the homoskedastic and with the heteroskedastic error model, and for each pair (p, r) and each value of n the exact covariance matrix was computed according to Equation (23). Figure 4 shows the estimated standard deviation (rms) of both fits, along with the exact values, for Φ , ϕ and κ , with $p = 0.8$, $r = 0.2$ in the range $3 \leq n \leq 20$. The standard deviations of the heteroskedastic fit are not only smaller, but also fall at a faster rate than the ones of the homoskedastic fit, i.e., faster than $1/\sqrt{n}$. The agreement between the estimated and the exact standard deviations is again excellent.

Figures 5–7 show the relative standard deviation for the three track parameters and all 20 combinations of p and r . Because of the large correlation between ϕ and κ , Figures 6 and 7 are virtually identical. The same trend as before can be observed: the ratio gets smaller with decreasing p and with decreasing r . The largest improvement is again seen for $p = 0.6$, $r = 0.1$, where the probability of very precise measurements ($r = 0.1$) is the largest ($1 - p = 0.4$). Compared to the results of the straight-line fits, the gain in precision is somewhat smaller.

4.3. Simulation Study with Circular Tracks and Multiple Scattering

The third study investigates the effect of multiple Coulomb scattering (MCS) on the results presented in the previous subsection. To this end, every layer of the toy detector is assumed to have a thickness of 1% of a radiation length. The simulation and the reconstruction of the tracks is modified accordingly, see for example [17].

10^5 tracks were generated for each of four values of the transverse momentum, i.e., $p_T = 10, 20, 50, 100$ GeV/ c and for each of the 20 combinations of p and $r = \sigma_0/\sigma_1$ as in Section 4.2. The dip angle of the tracks was set to $\lambda = \pi/4$. The track position in the bending plane was recorded in each cylinder, with a position error generated from the Gaussian mixtures in Table 1 plus a random deviation due to MCS. The number n of planes varied from 3 to 20.

It is to be expected that the benefit provided by the heteroskedastic regression is diminished by the introduction of MCS. This is confirmed by the results in Figures 8–11, which show the relative standard deviation of the curvature κ for the four values of p_T and the 20 combinations of p and r . Although the gain in precision of the heteroskedastic estimator is rather poor for the lowest value of p_T , see Figure 8, it steadily improves with increasing p_T and approaches the optimal level for the highest value of p_T , see Figures 7 and 11.

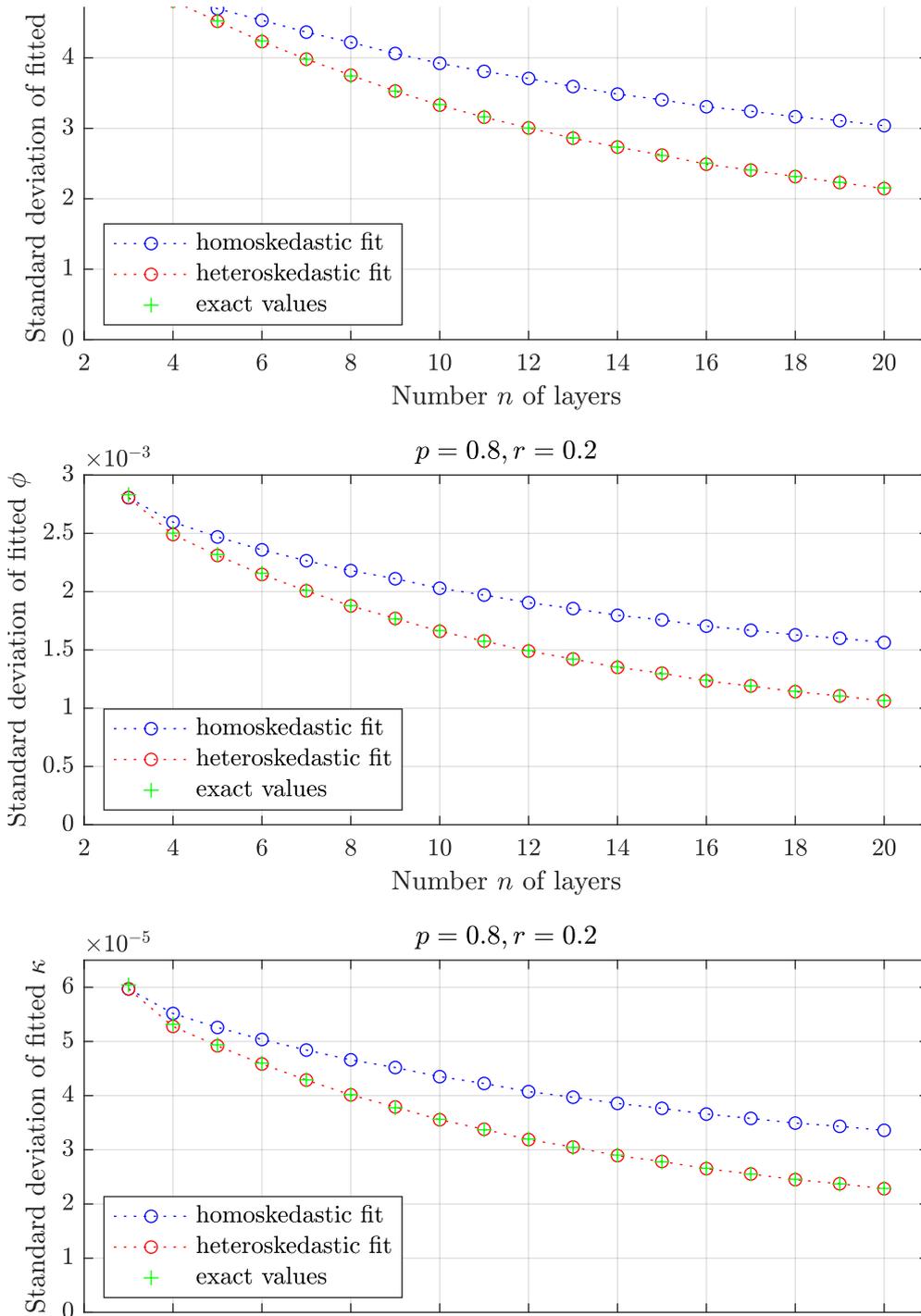


Figure 4. **Top:** estimated rms and exact standard deviation of the fitted angle Φ ; **center:** estimated rms and exact standard deviation of the fitted angle ϕ ; **bottom:** estimated rms and exact standard deviation of the fitted curvature κ , for $p = 0.8, r = 0.2$ in the range $3 \leq n \leq 20$.

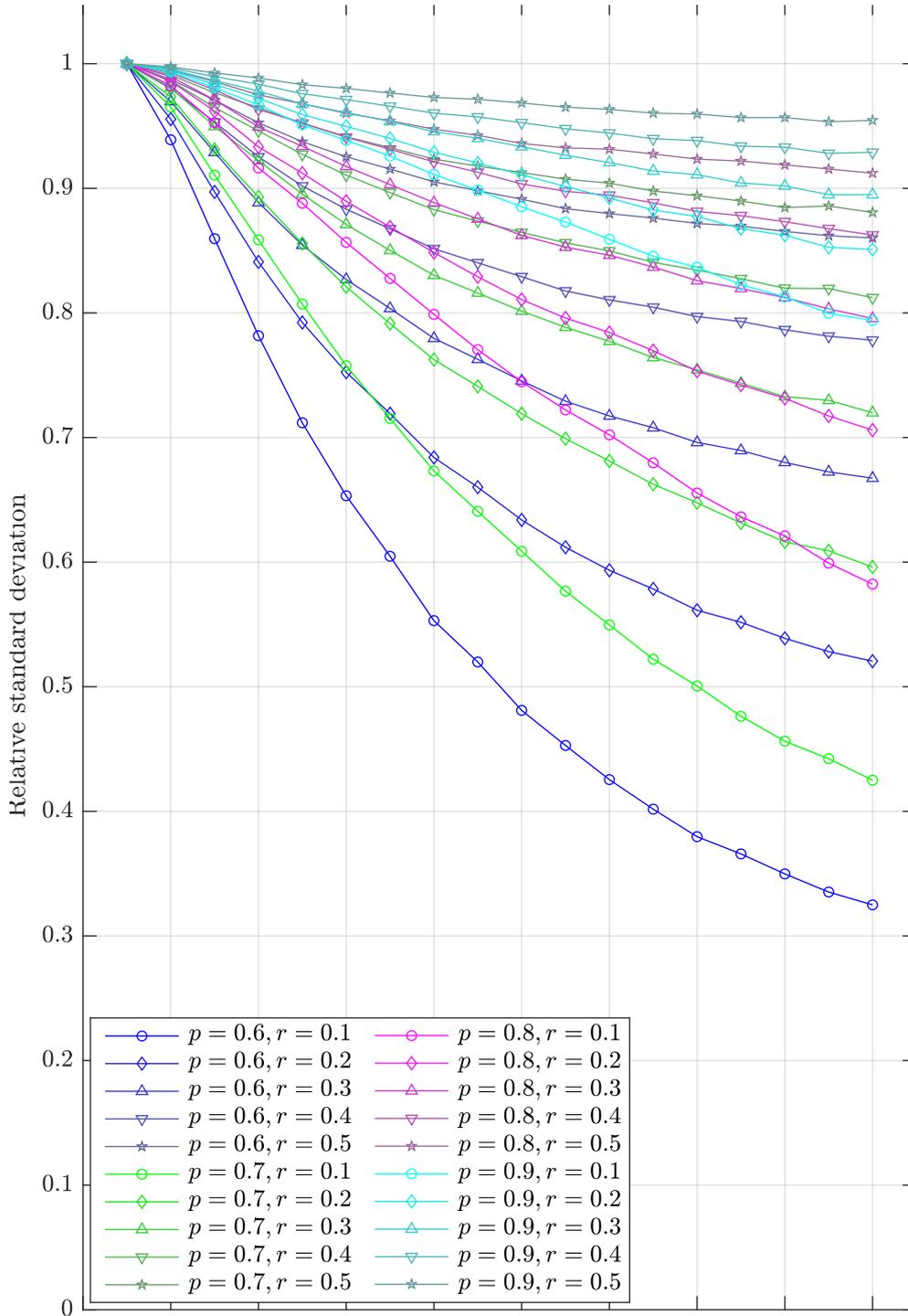


Figure 5. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the fitted angle Φ for all 20 combinations of p and r in the range $3 \leq n \leq 20$.

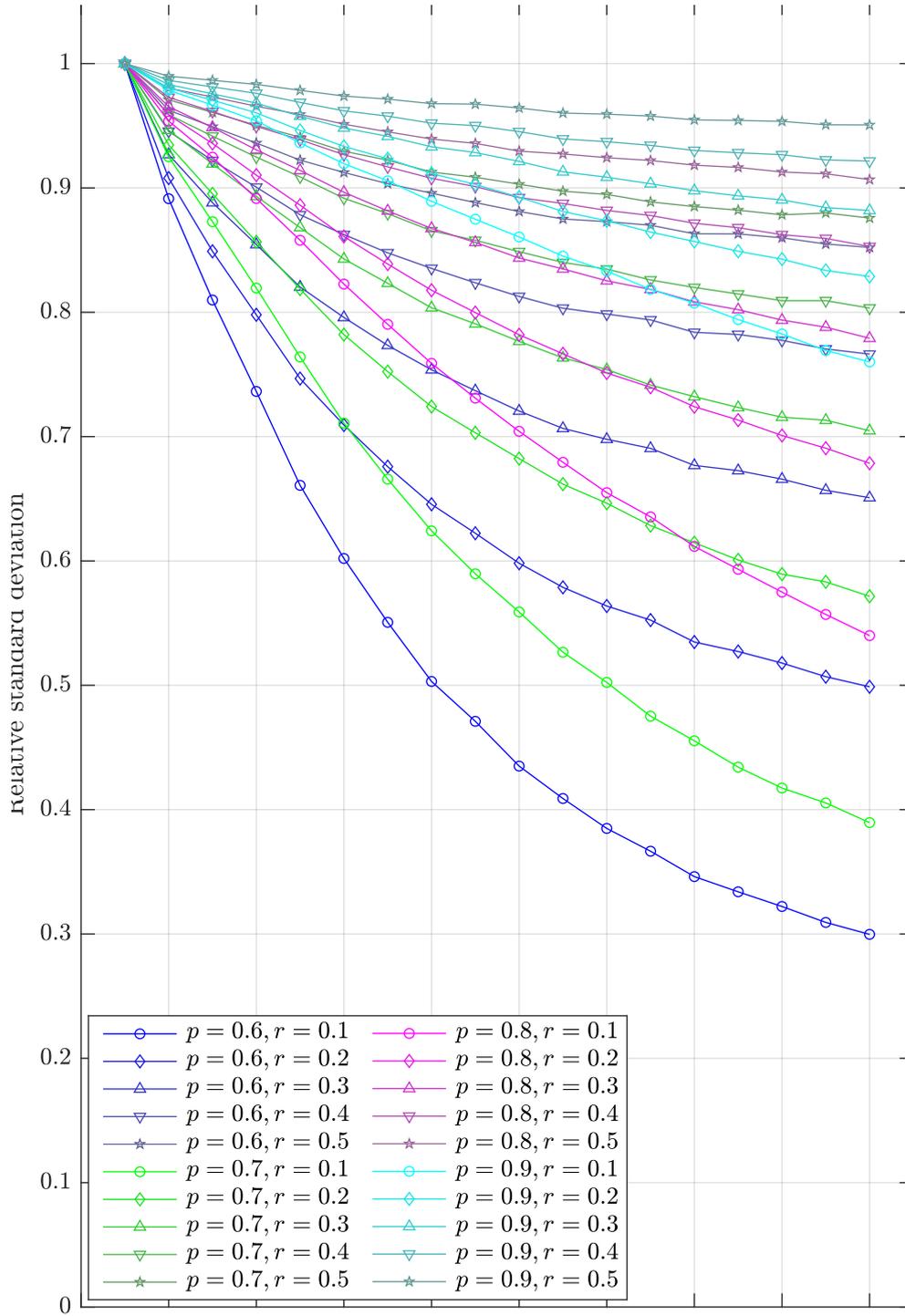


Figure 6. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the fitted angle ϕ for all 20 combinations of p and r in the range $3 \leq n \leq 20$.

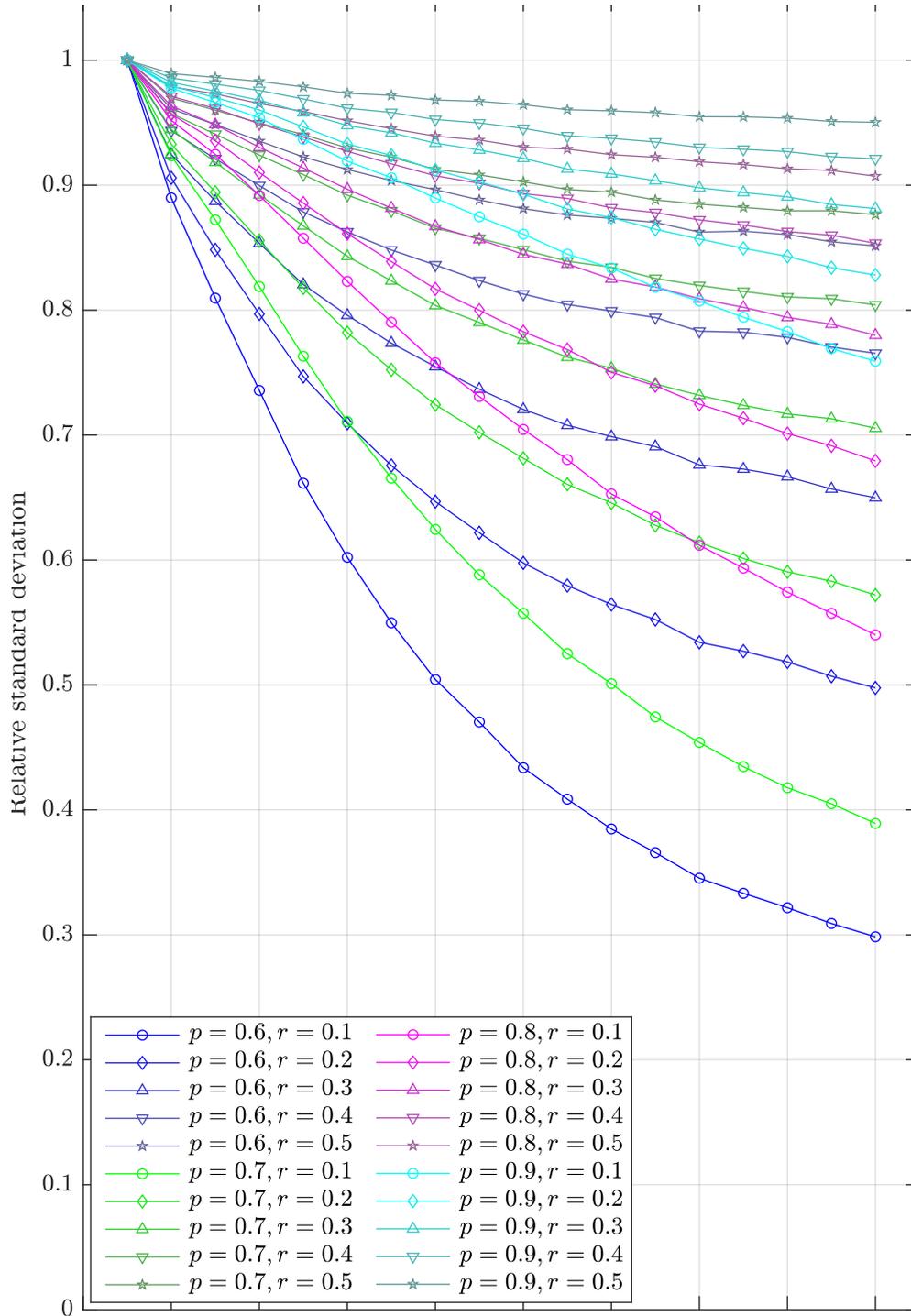


Figure 7. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the fitted curvature κ for all 20 combinations of p and r in the range $3 \leq n \leq 20$.

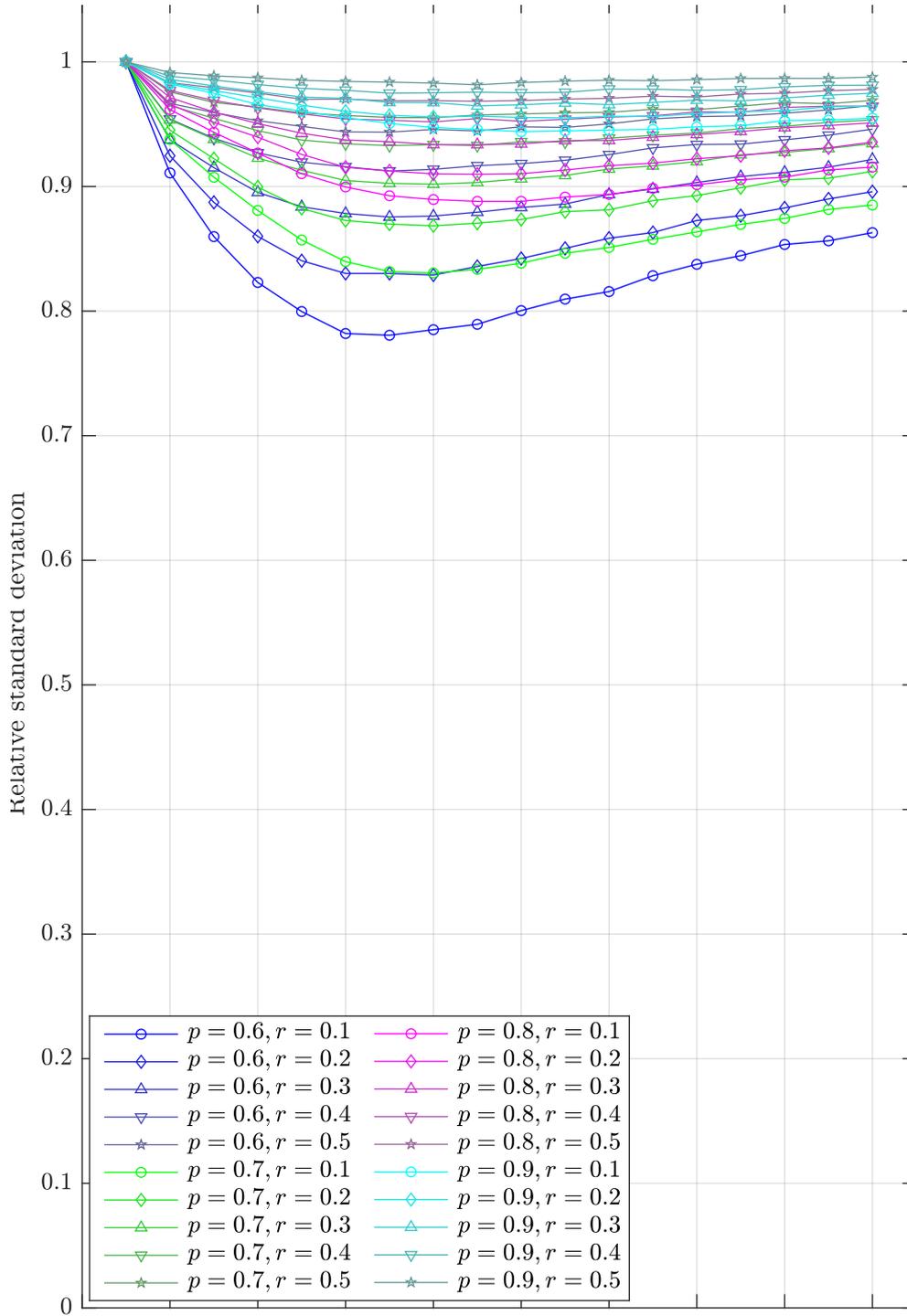


Figure 8. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the curvature κ for $p_T = 10, \text{GeV}/c$, $\lambda = \pi/4$ and all 20 combinations of p and r in the range $3 \leq n \leq 20$.

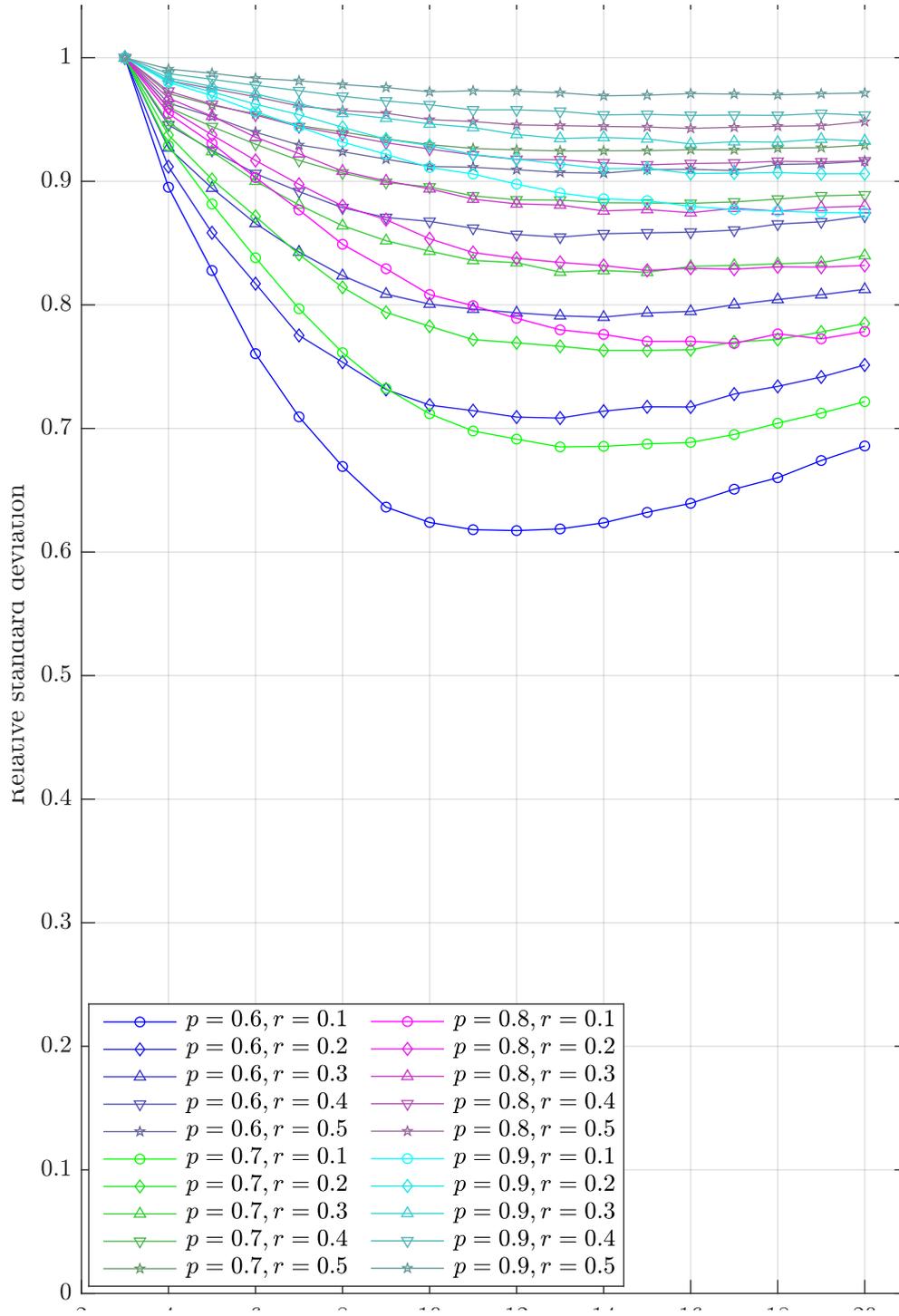


Figure 9. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the curvature κ for $p_T = 20, \text{GeV}/c, \lambda = \pi/4$ and all 20 combinations of p and r in the range $3 \leq n \leq 20$.

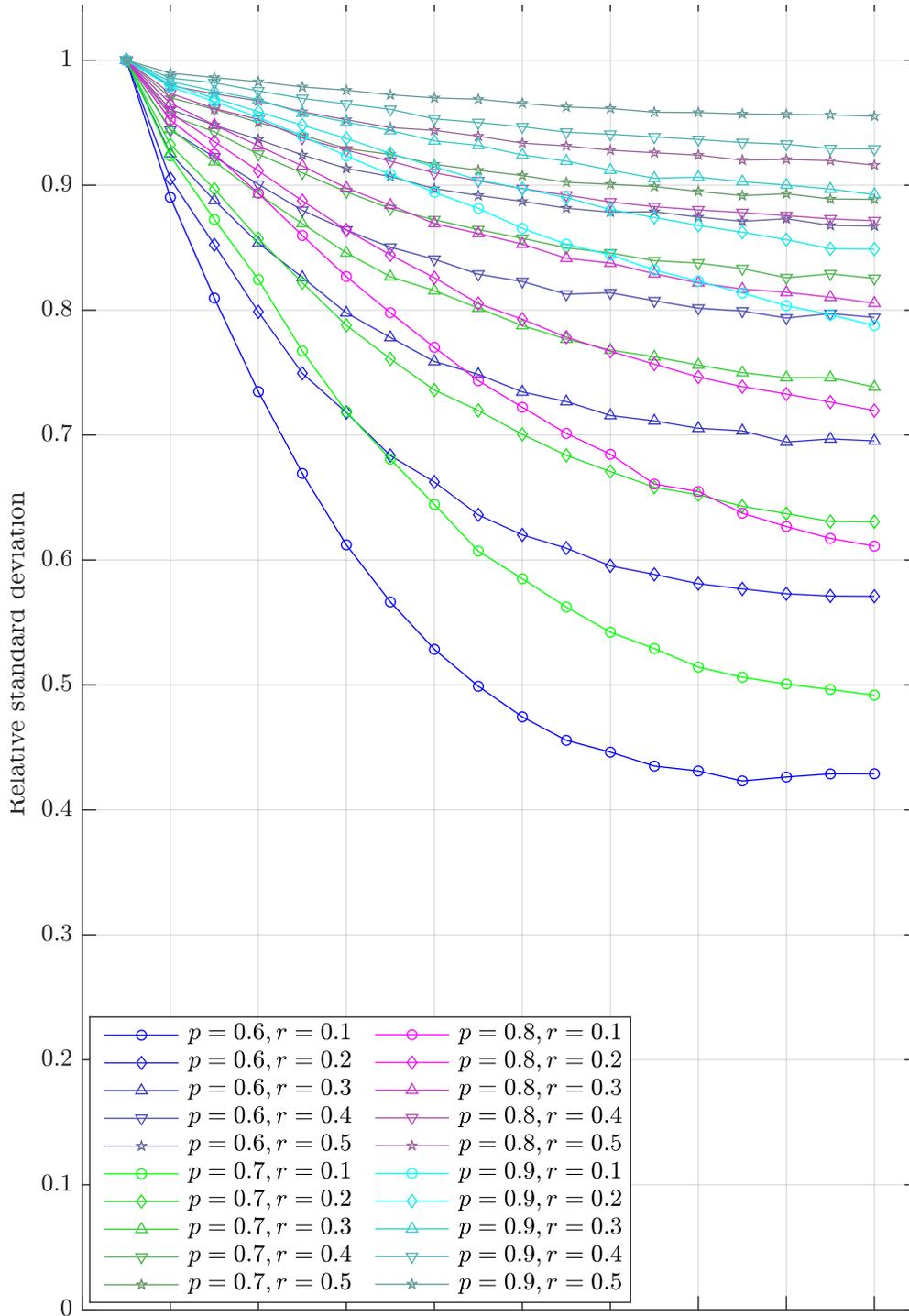


Figure 10. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the curvature κ for $p_T = 50, \text{GeV}/c, \lambda = \pi/4$ and all 20 combinations of p and r in the range $3 \leq n \leq 20$.

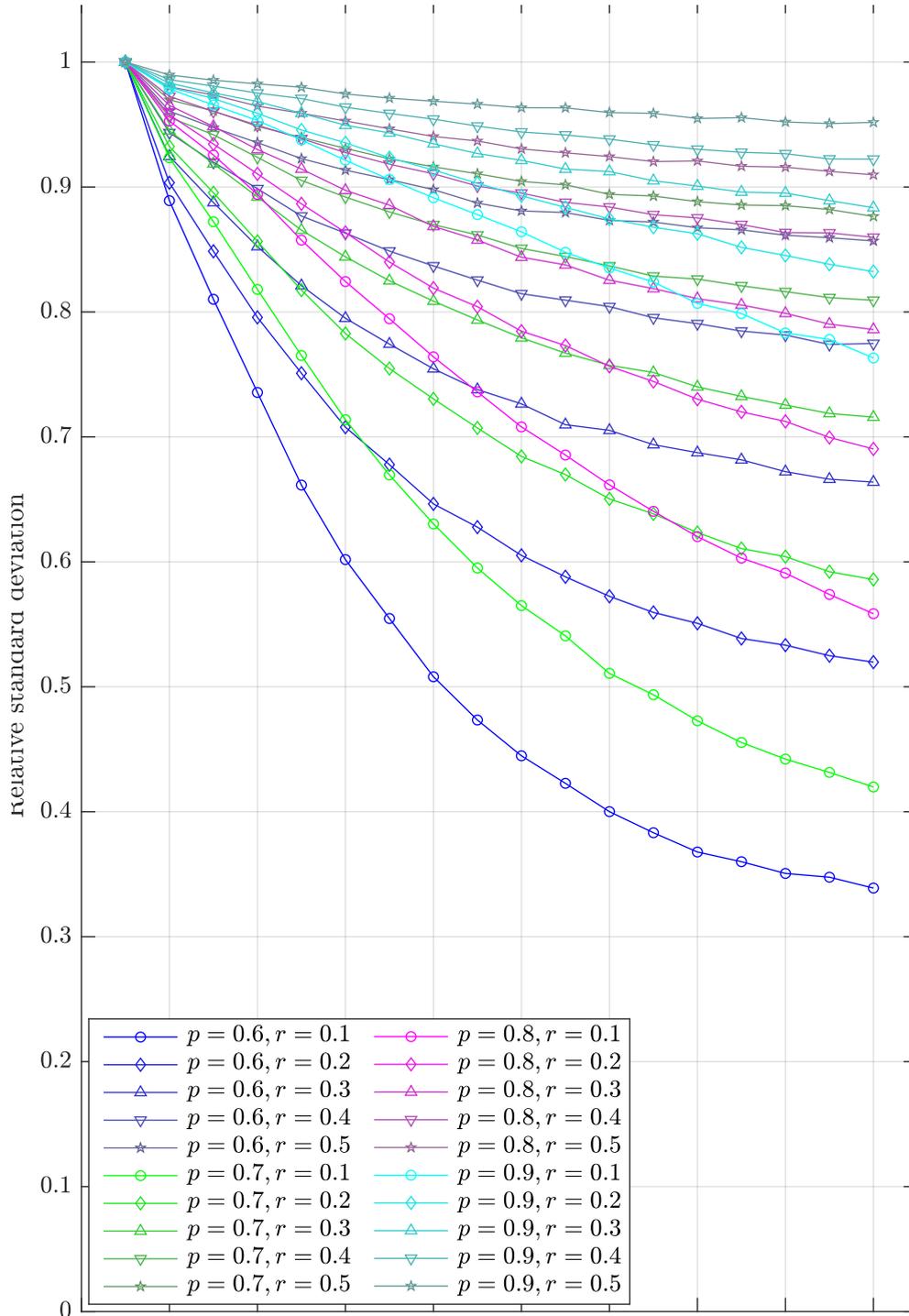


Figure 11. Relative standard deviation (heteroskedastic rms over homoskedastic rms) of the curvature κ for $p_T = 100, \text{GeV}/c$, $\lambda = \pi/4$ and all 20 combinations of p and r in the range $3 \leq n \leq 20$.

5. Summary and Conclusions

We have evaluated the gain in precision that can be achieved by GMR in track fitting using a mixture model of position errors rather than a simple Gaussian model with an average resolution. The results, both with straight and with circular tracks, show the expected behavior: the gain rises if the mixture weight (proportion) of the narrow component becomes larger, and it rises when the ratio of the width of the narrow component and the width of the wide component becomes smaller.

Unsurprisingly, the gain is negatively affected by MCS, depending on the track momentum and on the amount of material.

It is also instructive to look at the results in the light of the findings reported in [12–14], in particular the claim that the heteroskedastic regression, which is in fact the GMR described here, achieves a linear growth of the resolution with the number n of detecting layers, instead of the usual \sqrt{n} behavior [13]. A glance at Figure 3 shows that in fact the heteroskedastic standard deviations initially fall faster than the homoskedastic ones, i.e., faster than $1/\sqrt{n}$. Eventually, however, the relative standard deviations level out at a constant ratio, showing that asymptotically both standard deviations obey the expected $1/\sqrt{n}$ mode of falling with n . The observation reported in [13] is clearly a small-sample effect, which—in contrast to most other cases—is very significant in the mixture with $p = 0.8, r = 0.1$, the one assumed in [13]. Fortunately, it is in the range of n that is relevant for realistic tracking detectors, i.e., somewhere between 5 and 20 for a silicon tracker. It seems therefore absolutely worthwhile to optimize the local reconstruction and error calibration of the position measurements as far as possible.

Funding: This research received no external funding.

Acknowledgments: I thank Gregorio Landi for interesting and instructive discussions. I also thank the anonymous reviewers for valuable comments and corrections, and in particular for suggesting to investigate the effects of multiple Coulomb scattering.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GM	Gaussian Mixture
GMM	Gaussian Mixture Model
GMR	Gaussian Mixture Regression
LGRM	Linear Gaussian Regression Model
LS	Least-Squares
MCS	Multiple Coulomb Scattering
PDF	Probability Density Function

References

- Frühwirth, R. Track fitting with non-Gaussian noise. *Comput. Phys. Commun.* **1997**, *100*, 1. [[CrossRef](#)]
- Frühwirth, R.; Strandlie, A. Track fitting with ambiguities and noise: A study of elastic tracking and nonlinear filters. *Comput. Phys. Commun.* **1999**, *120*, 197. [[CrossRef](#)]
- Frühwirth, R.; Regler, M. On the quantitative modelling of core and tails of multiple scattering by Gaussian mixtures. *Nucl. Instrum. Meth. Phys. Res. A* **2001**, *456*, 369. [[CrossRef](#)]
- Frühwirth, R.; Liendl, M. Mixture models of multiple scattering: computation and simulation. *Comput. Phys. Commun.* **2001**, *141*, 230. [[CrossRef](#)]
- Frühwirth, R. A Gaussian-mixture approximation of the Bethe-Heitler model of electron energy loss by bremsstrahlung. *Comput. Phys. Commun.* **2003**, *154*, 131. [[CrossRef](#)]
- Adam, W.; Frühwirth, R.; Strandlie, A.; Todorov, T. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *J. Phys. G Nucl. Part. Phys.* **2005**, *31*, N9. [[CrossRef](#)]
- Strandlie, A.; Frühwirth, R. Reconstruction of charged tracks in the presence of large amounts of background and noise. *Nucl. Instrum. Meth. Phys. Res. A* **2006**, *566*, 157. [[CrossRef](#)]
- The ATLAS Collaboration. *Improved Electron Reconstruction in ATLAS Using the Gaussian Sum Filter-Based Model for Bremsstrahlung*; Technical Report ATLAS-CONF-2012-047; CERN: Geneva, Switzerland, 2012.
- Strandlie, A.; Frühwirth, R. Track and vertex reconstruction: From classical to adaptive methods. *Rev. Mod. Phys.* **2010**, *82*, 1419. [[CrossRef](#)]
- Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B* **1977**, *39*, 1.

11. Borman, S. *The Expectation Maximization Algorithm—A Short Tutorial*; Technical report; University of Utah: Salt Lake City, UT, USA, 2004. Available online: tinyurl.com/BormanEM (accessed on 30 August 2020).
12. Landi, G.; Landi, G. Optimizing Momentum Resolution with a New Fitting Method for Silicon-Strip Detectors. *Instruments* **2018**, *2*, 22. [[CrossRef](#)]
13. Landi, G.; Landi, G. Beyond the \sqrt{N} Limit of the Least Squares Resolution and the Lucky Model. 2018. Available online: <http://xxx.lanl.gov/abs/1808.06708> (accessed on 20 August 2020).
14. Landi, G.; Landi, G. The Cramer–Rao Inequality to Improve the Resolution of the Least-Squares Method in Track Fitting. *Instruments* **2020**, *4*, 2. [[CrossRef](#)]
15. Kendall, M.; Stuart, A. *The Advanced Theory of Statistics*, 3rd ed.; Charles Griffin: London, UK, 1973; Volume 2.
16. Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **1894**, 185.
17. Brondolin, E.; Frühwirth, R.; Strandlie, A. Pattern recognition and reconstruction. In *Particle Physics Reference Library Volume 2: Detectors for Particles and Radiation*; Fabjan, C.; Schopper, H., Eds.; Springer International Publishing: Cham, Switzerland, 2020.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).