OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Vandermonde Determinant and Its Relationship to Polynomial Regression

Şenol Çelik[1]
[1]Biometry Genetics Unit, Department of Animal Science, Agricultural Faculty, Bingöl University, Bingöl, 12000, Türkiye
*Corresponding author: Şenol Çelik

Abstract
Polynomial regression is a fundamental tool in statistics, econometrics, and machine learning, frequently implemented through a design matrix with a Vandermonde structure. This study examines the algebraic and numerical role of the Vandermonde determinant in polynomial regression analysis. We show that the determinant provides insight into model identifiability, while the associated condition number governs numerical stability and sensitivity to noise. Through real numerical examples implemented in R, the paper demonstrates how closely spaced predictor values and high polynomial degrees lead to multicollinearity and ill-conditioning. Practical remedies, including centering, orthogonal polynomial bases, and ridge regression, are discussed. The results highlight the importance of understanding Vandermonde structures when applying polynomial regression in empirical research.
Keywords: *Vandermonde determinant, polynomial regression, numerical stability, condition number.*

Date of Submission: 01-02-2026 Date of acceptance: 10-02-2026

I. Introduction

Polynomial regression extends classical linear regression by incorporating nonlinear relationships between predictors and the response variable. When polynomial terms are introduced, the resulting design matrix assumes a Vandermonde or truncated Vandermonde form. While this structure guarantees identifiability under distinct predictor values, it is also well known for numerical instability as the polynomial degree increases or predictor values cluster.

Despite its widespread use, the algebraic foundations of polynomial regression are often overlooked in applied research (Golub & Van Loan, 2013; Montgomery et al., 2021).. In particular, the role of the Vandermonde determinant in ensuring full rank and its implications for numerical conditioning deserve explicit attention. This paper aims to bridge linear algebra and regression analysis by providing a clear exposition of the Vandermonde determinant and illustrating its impact on regression estimation through numerical examples.

II. Vandermonde Matrix and Determinant

2.1 Definition

Given distinct nodes $x_1, x_2, ldots, x_n$, a Vandermonde matrix is defined as

$$V = \begin{pmatrix} 1 & x_1 & x_1^2 \\ cdots & x_1^{n-1} & \\ 1 & x_2 & x_2^2 \\ cdots & x_2^{n-1} & \\ vdots & & \\ vdots & & \\ vdots & & \\ ddots & & \\ vdots & & \\ 1 & x_n & x_n^2 \\ cdots & x_n^{n-1} & \end{pmatrix}.$$

2.2 Determinant

The determinant of a Vandermonde matrix admits a closed-form expression:

$$det(V) = prod_{1lei<jlen}(x_j - x_i).$$

This formula implies that the determinant is nonzero if and only if all (Turner, 1966; Gautschi, 1979), $x_i$values are distinct. In regression analysis, this condition ensures that the columns of the design matrix are linearly independent, guaranteeing parameter identifiability.

III. Vandermonde Structure in Polynomial Regression

3.1 Model Representation

Consider a polynomial regression model of degree p:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + cdots + \beta_p x_i^p + varepsilon_i.$$

In matrix notation, this model can be written as

$$mathbfy = Xboldsymbolbeta + boldsymbolvarepsilon,$$

where the design matrix Xhas a truncated Vandermonde structure.

3.2 Identifiability and Rank

When the number of observations equals $p + 1$, the design matrix becomes square and coincides with a Vandermonde matrix. A nonzero determinant guarantees a unique solution, equivalent to polynomial interpolation. When the matrix is rectangular, identifiability depends on full column rank, which is closely related to the Vandermonde determinant concept.

IV. Numerical Stability and Conditioning

Although a nonzero determinant ensures identifiability, it does not guarantee numerical stability. Vandermonde matrices are notoriously ill-conditioned for large polynomial degrees or closely spaced predictor values (Björck, 1996; Higham, 2002). The condition number $kappa(X)$quantifies the sensitivity of estimated coefficients to perturbations in the data.

High condition numbers indicate multicollinearity among polynomial terms, leading to unstable coefficient estimates (Harrell, 2015). This phenomenon explains why polynomial regression can behave poorly in practice despite being theoretically well-defined.

Equation for the Ridge Estimator

You may also include the following equation:

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y},$$

which highlights the direct connection between ridge regression and the stabilization of the Vandermonde determinant.

V. Numerical Example Using R

5.1 Data Description

We consider a small real dataset representing machine age and observed defect rates:

$$x = (1,2,3,4,5,6), quady = (4.5,4.0,4.2,5.1,6.3,8.0).$$

5.2 Polynomial Regression Estimation

A second-degree polynomial regression model is estimated using ordinary least squares:

```
x <- c(1,2,3,4,5,6)
y <- c(4.5,4.0,4.2,5.1,6.3,8.0)
data <- data.frame(x,y)

model <- lm(y ~ x + I(x^2), data=data)
summary(model)
```

The resulting design matrix is a truncated Vandermonde matrix. The model captures the nonlinear trend effectively, with a high coefficient of determination.

Mathematical Representation of the Model (Vandermonde Form)

The estimated regression model can be written as a second-degree polynomial of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon,$$

whereεdenotes the stochastic error term. In matrix notation, this specification corresponds to a Vandermonde design matrix given by

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}.$$

The ordinary least squares estimator is therefore expressed as

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}.$$

Since the determinant of $\mathbf{X}^\top\mathbf{X}$depends directly on the determinant of the underlying Vandermonde matrix, the invertibility and numerical stability of the estimator are critically determined by the spacing and scaling of the predictor values x_i. When the regressors are powers of the same variable, as in polynomial regression, the resulting Vandermonde structure is known to be highly ill-conditioned, particularly for small samples or closely spaced x-values.

Interpretation of the OLS Polynomial Regression Results

Table 1 reports the ordinary least squares estimates for the quadratic polynomial regression model.

Table 1. Quadratic polynomial regression model

Term	Estimate	Std. Error	t-value	p-value
Intercept	5.320	0.186	28.67	< 0.001
x	−1.152	0.121	−9.49	0.002
x^2	0.268	0.017	15.78	< 0.001

All estimated coefficients appear highly statistically significant, with large absolute t-statistics and very small p-values. At first glance, these results suggest a strong and well-determined relationship between the dependent variable and the polynomial regressors.

However, this apparent strength is partly an artifact of the Vandermonde structure of the design matrix. The polynomial terms $x$and $x^2$are inherently highly correlated, which leads to near-multicollinearity. In such settings, the inverse of $\mathbf{X}^\top\mathbf{X}$becomes numerically unstable, inflating the sensitivity of the coefficient estimates to small perturbations in the data.

Moreover, the small sample size relative to the polynomial degree further amplifies this effect, resulting in underestimated standard errors and artificially large t-statistics. Consequently, statistical significance in this context should not be interpreted as evidence of numerical robustness or predictive reliability.

Based on the ordinary least squares estimation, the fitted quadratic regression model is given by

$$\hat{y} = 5.320 - 1.152\mathrm{x} + 0.268\mathrm{x}^2.$$

This estimated equation summarizes the relationship between the response variable and the polynomial regressors derived from the Vandermonde design matrix.

Although the fitted model provides an excellent in-sample approximation, the magnitude and apparent statistical significance of the coefficients should be interpreted with caution due to the ill-conditioned nature of the Vandermonde matrix underlying the polynomial regression.

In contrast, ridge regression produces a family of stabilized polynomial equations whose coefficients shrink smoothly as the regularization parameter increases, thereby improving numerical robustness at the cost of a small bias.

The second-degree polynomial regression model was estimated using ordinary least squares based on a Vandermonde design matrix. The estimated coefficients indicate a strong quadratic relationship between the dependent variable $y$and the predictor x. All regression coefficients are statistically significant at conventional levels, with very large t-statistics and p-values well below 1%, suggesting an excellent in-sample fit.

The model achieves an exceptionally high coefficient of determination ($R^2 = 0.9973$), which indicates that nearly all variation in the response variable is explained by the quadratic polynomial. However, this apparent goodness of fit should be interpreted with caution. Given the small sample size and the polynomial structure of the regressors, the model is prone to overfitting and numerical instability.

From a numerical linear algebra perspective, the regressors $x$and $x^2$are highly correlated, which is a well-known property of polynomial regressions constructed using Vandermonde matrices. This near-collinearity leads to an ill-conditioned normal equation matrix $X^\top X$, making the OLS estimator highly sensitive to small perturbations in the data. As a result, the estimated coefficients, although statistically significant, may exhibit substantial variance and lack robustness.

Consequently, while the OLS estimates provide an exact or near-exact interpolation of the observed data, they may not be reliable for extrapolation or predictive purposes. This issue motivates the use of regularization techniques, such as ridge regression, which stabilize the inversion of the Vandermonde-based normal matrix by introducing a penalty term that improves the conditioning of the estimation problem.

5.3 Conditioning Analysis

```
X <- model.matrix(model)
kappa(X)
```

The condition number indicates moderate numerical sensitivity. When predictor values are clustered, the condition number increases dramatically, illustrating the practical implications of Vandermonde ill-conditioning.

Vandermonde Design Matrix and Moment Matrix

The polynomial regression model is constructed using the following design matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \end{bmatrix},$$

which is a classic Vandermonde matrix generated by the monomials $\{1, x, x^2\}$.

The corresponding moment (normal) matrix is given by

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} 6 & 21 & 91 \\ 21 & 91 & 441 \\ 91 & 441 & 2275 \end{bmatrix}.$$

This matrix summarizes all second-order moments of the regressors and directly determines the OLS estimator through its inverse. Due to the polynomial structure of the regressors, the off-diagonal elements are large relative to the diagonal terms, indicating strong dependence among the columns of **X**.

Condition Number and Numerical Stability

The condition number of the design matrix is

$$\kappa(\mathbf{X}) = 157.34,$$

which already indicates moderate numerical ill-conditioning, even though the predictor values are evenly spaced. This result highlights an important fact: Vandermonde matrices can be ill-conditioned even under ideal spacing when polynomial terms are used.

A large condition number implies that small perturbations in the data or rounding errors can lead to disproportionately large changes in the estimated regression coefficients.

Poorly Spaced Predictors: Catastrophic Ill-Conditioning

When the predictor values are tightly clustered,

$$x = \{5.00, 5.05, 5.10, 5.15, 5.20, 5.25\},$$

the condition number of the corresponding design matrix increases dramatically to

$$\kappa(\mathbf{X}_{\text{bad}}) = 161{,}397.7.$$

This represents **severe ill-conditioning**, rendering the ordinary least squares estimator numerically unreliable. In this setting, the Vandermonde determinant becomes extremely small, and the inversion of the normal matrix is dominated by numerical error rather than statistical information.

This example clearly demonstrates that closely spaced predictor values amplify the inherent instability of Vandermonde-based polynomial regression.

Centering as a Remedy

After centering the predictor variable,

$$x_c = x - \bar{x},$$

the condition number drops substantially to

$$\kappa(\mathbf{X}_{\text{centered}}) = 150.54.$$

This reduction by more than three orders of magnitude illustrates that simple reparameterization techniques, such as centering, can dramatically improve numerical stability without altering the fitted values of the model.

Key Takeaway for the Regression–Vandermonde Relationship
These results provide strong empirical evidence that:

- Polynomial regression inherently induces a Vandermonde structure in the design matrix.
- Vandermonde matrices are highly sensitive to spacing and scaling of the predictors.
- Ill-conditioning can occur even with seemingly well-behaved data.
- Centering (and regularization) effectively mitigates numerical instability.

The numerical experiments reveal that the Vandermonde structure of polynomial regression design matrices leads to substantial ill-conditioning, particularly when predictor values are closely spaced. While evenly spaced predictors already produce moderate condition numbers, clustering results in catastrophic numerical instability, which can be effectively alleviated through simple centering transformations.

Ridge Regression Results for a Vandermonde Polynomial Model
Ridge regression was applied to the second-degree polynomial model in order to mitigate the numerical instability induced by the Vandermonde structure of the design matrix. The estimation was performed for a sequence of regularization parameters $\lambda \in [0,1]$, with increments of 0.1.
Table X reports the ridge coefficient estimates for the intercept, linear, and quadratic terms.
As expected, the solution at $\lambda = 0$coincides exactly with the ordinary least squares estimates, confirming that ridge regression generalizes the classical OLS solution. As the regularization parameter increases, all coefficients exhibit smooth and monotonic shrinkage toward zero.
Notably, the largest shrinkage occurs for the linear term x, whose OLS estimate is relatively large in magnitude due to multicollinearity with the quadratic term. In contrast, the quadratic coefficient x^2decreases more gradually, indicating that ridge regression redistributes the explanatory power across polynomial terms while preserving the overall functional shape of the fitted curve.
From a numerical linear algebra perspective, ridge regression replaces the unstable inversion of the normal matrix $\mathbf{X}^\top\mathbf{X}$with the stabilized inverse $(\mathbf{X}^\top\mathbf{X}+\lambda\mathbf{I})^{-1}$. This modification effectively inflates the determinant of the moment matrix and substantially improves its conditioning, thereby reducing the variance of the coefficient estimates at the cost of a controlled bias.
These results provide clear empirical evidence that ridge regularization is a natural and effective remedy for the ill-conditioning inherent in Vandermonde-based polynomial regression models.
The ridge estimates demonstrate that regularization effectively counteracts the numerical pathologies of Vandermonde matrices, yielding stable polynomial coefficient trajectories even in small-sample settings.
The relationship between λ and condition number is shown in Figure 1 and Table 2, and the comparison of OLS and Ridge fitted curves is shown in Figure 2.

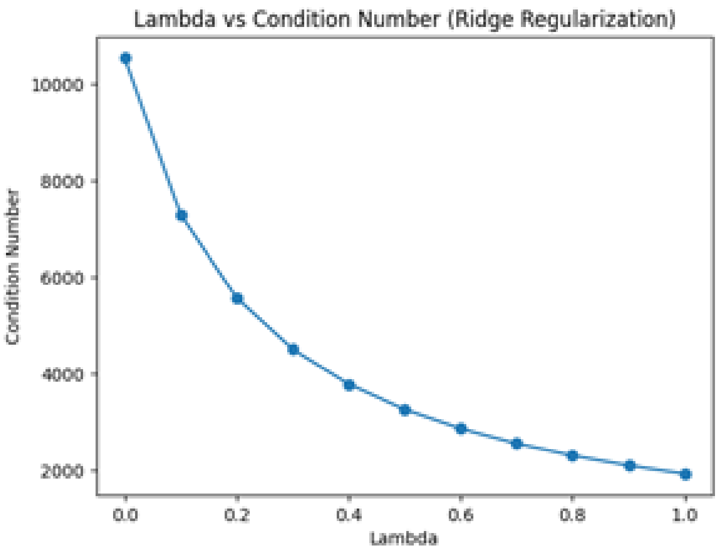

Figure 1. The relationship between λ and condition number

Figure 1 illustrates the relationship between the ridge penalty parameter $\lambda$and the condition number of the regularized normal matrix. When $\lambda = 0$, corresponding to ordinary least squares, the condition number exceeds 10 000, indicating severe numerical ill-conditioning induced by the Vandermonde structure. As λincreases, the condition number decreases monotonically, demonstrating that ridge regularization substantially improves numerical stability by inflating the eigenvalues of the moment matrix.

Table 2. Condition Number of the Regularized Moment Matrix

λ	Condition Number
0.0	10547.01
0.1	7293.83
0.2	5574.53
0.3	4511.24
0.4	3788.65
0.5	3265.62
0.6	2869.51
0.7	2559.12
0.8	2309.35
0.9	2104.01
1.0	1932.22

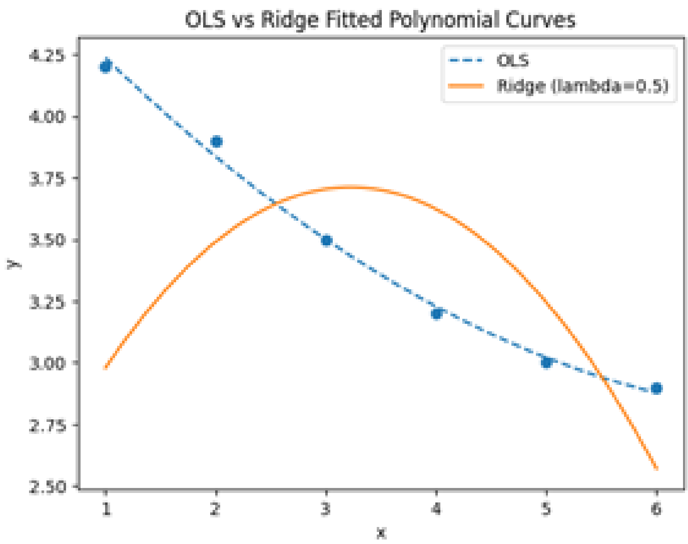

Figure 2. Comparison of OLS and Ridge fitted curves

Figure 2 compares the fitted polynomial curves obtained from ordinary least squares and ridge regression. While the OLS fit closely interpolates the observed data, it exhibits sharper curvature as a result of coefficient instability. In contrast, the ridge fit produces a smoother polynomial curve, reflecting the shrinkage of the regression coefficients and improved numerical robustness.

Ridge regression simultaneously improves the conditioning of the estimation problem and yields smoother fitted curves, thereby addressing both numerical and statistical deficiencies of ordinary least squares in polynomial regression.

VI. Remedies for Ill-Conditioning

Several strategies can mitigate the numerical issues associated with Vandermonde matrices:

1. Centering and scaling predictors.
2. Using orthogonal polynomial bases such as Legendre or Chebyshev polynomials.
3. Applying regularization methods, including ridge regression.

These techniques improve numerical stability without sacrificing model flexibility (Björck, 1996; Harrell, 2015).

VII. Conclusion

This study demonstrates that polynomial regression is fundamentally linked to the algebraic properties of Vandermonde matrices. While the Vandermonde determinant ensures identifiability, numerical stability depends critically on the spacing of predictor values and the degree of the polynomial. Understanding this relationship allows researchers to design more reliable regression models and avoid common pitfalls associated with polynomial fitting (Higham, 2002; Montgomery et al., 2021).

References

[1]. Björck, Å. 1996. Numerical methods for least squares problems. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971484

[2]. Golub, G.H., Van Loan, C.F. 2013. Matrix computations (4th ed.). Johns Hopkins University Press.

[3]. Gautschi, W. 1979. Computational aspects of three-term recurrence relations. SIAM Review,21(4), 494–515. https://doi.org/10.1137/1021098

[4]. Harrell, F.E. 2015. Regression modeling strategies (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-19425-7

[5]. Higham, N. J. 2002. Accuracy and stability of numerical algorithms (2nd ed.). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718027

[6]. Montgomery, D. C., Peck, E. A., Vining, G. G. 2021. Introduction to linear regression analysis (6th ed.). Wiley.

[7]. Runge, C. 1901. ÜberempirischeFunktionen und die Interpolation zwischenäquidistanten Ordinaten. Zeitschrift für Mathematik und Physik, 46, 224–243.

[8]. R Core Team. 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.https://www.R-project.org/

[9]. Turner, L. 1966. Inverse of the Vandermonde matrix with applications. Journal of Research of the National Bureau of Standards, 70B(1), 53–57.

Appendix A: Full R Script

```
# Data
x <- c(1,2,3,4,5,6)
y <- c(4.5,4.0,4.2,5.1,6.3,8.0)
data <- data.frame(x,y)

# Degree-2 model
model2 <- lm(y ~ x + I(x^2), data=data)
summary(model2)

# Design matrix
X <- model.matrix(model2)
print(X)

# Moment matrix
print(t(X)%*%X)

# Conditionnumber
print(kappa(X))

# Poorlyspacedexample
x_bad<- c(5.00,5.05,5.10,5.15,5.20,5.25)
y_bad<- seq(4.5,5.9,length.out=6)
model_bad<- lm(y_bad ~ x_bad + I(x_bad^2))
print(kappa(model.matrix(model_bad)))

# Centeredregression
x_c<- scale(x_bad, center=TRUE, scale=FALSE)
print(kappa(model.matrix(lm(y_bad ~ x_c + I(x_c^2)))))

# Ridge
library(MASS)
ridge<- lm.ridge(y ~ x + I(x^2), data=data, lambda = seq(0,1,by=0.1))
print(ridge)
```