Large Deviation Property of the Condition Number of a Rectangular Random Vandermonde-like Matrix

Yan Zeng

May 9, 2012

1 Formulation of the problem

We consider the following problem of polynomial interpolation. Suppose we observe two sequences, $(x_i)_{i=1}^n$ and $(y_i)_{i=1}^n$. If y depends on x via the equation $y = a_0 + a_1 x + a_2 x^2$, we need to solve the following system of linear equations to determine a_0, a_1 and a_2:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The $n \times 3$ matrix at the left side of the equation is a transposed rectangular Vandermonde matrix, so it is of full rank when x_i's are distinct, and the interpolation problem has a unique solution.

Although Vandermonde matrix is known to be numerically unstable (see Gautschi [3] for a survey), we have a better situation in the context of Monte Carlo simulation. Indeed, if we assume $(x_i)_{i=1}^n$ is a sequence of i.i.d. samples of some distribution function ν, then *with large probability, the condition number of the above rectangular Vandermonde matrix is small, and the probability of the exceptional set decays exponentially as n increases.*

Why should we care about such a problem? In derivative pricing, the Longstaff-Schwartz algorithm [2] relies on solving the above polynomial interpolation problem. If the result produced by their algorithm differs from benchmark (produced by, for example, embedded Monte Carlo simulation), there could be two possible causes: either quadratic polynomial is not the appropriate functional form, or the error is due to numerical instability. The large deviation property of the condition number allows us to focus on choosing the appropriate functional form.

2 Transformation of the problem

In section, we briefly explain the insight of Bass and Gröchenig [1], which inspires the discovery of large deviation property of the condition number.

Denote by $\mathcal{P}_2$ the space of polynomials of degree no more than 2. We want to find $p(x) \in \mathcal{P}_2$ such that $p(x_i) = y_i$ $(i = 1, \cdots, n)$.

Define $\mathbf{a} = (a_0, a_1, a_2)^T$, $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T$, $p(x; \mathbf{a}) = a_0 + a_1 x + a_2 x^2$, and

$$\mathcal{U} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

Let $\mathcal{T} = \mathcal{U}^T \mathcal{U}$, then $\kappa(\mathcal{U}) = \sqrt{\kappa(\mathcal{T})}$ and

$$\kappa(\mathcal{T}) = \frac{\max\{\text{eigenvalues of } \mathcal{T}\}}{\min\{\text{eigenvalues of } \mathcal{T}\}}$$

We note

$$\max\{\text{eigenvalues of } \mathcal{T}\} = \sup_{\mathbf{a}\in\mathbb{R}^3\setminus\{0\}} \frac{\langle \mathcal{T}\mathbf{a},\mathbf{a}\rangle}{\|\mathbf{a}\|_2^2} = \sup_{\mathbf{a}\in\mathbb{R}^3\setminus\{0\}} \frac{\langle \mathcal{U}\mathbf{a},\mathcal{U}\mathbf{a}\rangle}{\|\mathbf{a}\|_2^2} = \sup_{\mathbf{a}\in\mathbb{R}^3\setminus\{0\}} \frac{\sum_{i=1}^n |p(x_i;\mathbf{a})|^2}{\|\mathbf{a}\|_2^2}$$

and

$$\min\{\text{eigenvalues of } \mathcal{T}\} = \inf_{\mathbf{a}\in\mathbb{R}^3\setminus\{0\}} \frac{\langle \mathcal{T}\mathbf{a},\mathbf{a}\rangle}{\|\mathbf{a}\|_2^2} = \inf_{\mathbf{a}\in\mathbb{R}^3\setminus\{0\}} \frac{\langle \mathcal{U}\mathbf{a},\mathcal{U}\mathbf{a}\rangle}{\|\mathbf{a}\|_2^2} = \inf_{\mathbf{a}\in\mathbb{R}^3\setminus\{0\}} \frac{\sum_{i=1}^n |p(x_i;\mathbf{a})|^2}{\|\mathbf{a}\|_2^2}.$$

Therefore, the problem of estimating the condition number of the random matrix

$$\mathcal{U} = \begin{bmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ \vdots & \vdots & \vdots \\ 1 & X_n & X_n^2 \end{bmatrix}$$

is equivalent to estimating the range of the random variable

$$\frac{\sum_{i=1}^n |p(X_i;\mathbf{a})|^2}{\|\mathbf{a}\|_2^2}.$$

$\mathcal{P}_2$ is a finite dimensional space and $\|\mathbf{a}\|_2$ is a norm on $\mathcal{P}_2$. So it is equivalent to the following norm on $\mathcal{P}_2$:

$$\|p(\cdot;\mathbf{a})\|_{2,\nu} := \sqrt{E|p(X_1;\mathbf{a})|^2} = \left(\int |p(x;\mathbf{a})|^2 \nu(dx)\right)^{1/2}.$$

That is, there exist α, $\beta > 0$ such that

$$\alpha\|\mathbf{a}\|_2 \le \|p(\cdot;\mathbf{a})\|_{2,\nu} \le \beta\|\mathbf{a}\|_2, \ \forall p \in \mathcal{P}_2.$$

Define

$$Y_i(\mathbf{a}) = \frac{|p(X_i;\mathbf{a})|^2 - \|p(\cdot;\mathbf{a})\|_{2,\nu}^2}{\|p(\cdot;\mathbf{a})\|_{2,\nu}^2}$$

Then the inequality

$$\left|\sum_{i=1}^n Y_i(\mathbf{a})\right| < \lambda$$

is equivalent to

$$\frac{(-\lambda+n)\|p(\cdot;\mathbf{a})\|_{2,\nu}^2}{\|\mathbf{a}\|_2^2} < \frac{\sum_{i=1}^n |p(X_i;\mathbf{a})|^2}{\|\mathbf{a}\|_2^2} < \frac{(\lambda+n)\|p(\cdot;\mathbf{a})\|_{2,\nu}^2}{\|\mathbf{a}\|_2^2}$$

which gives the estimate of condition number

$$\kappa(\mathcal{T}) \le \frac{(\lambda+n)\beta}{(-\lambda+n)\alpha} = \frac{1+\frac{\lambda}{n}}{1-\frac{\lambda}{n}} \cdot \frac{\beta}{\alpha}$$

In summary, we have transformed the original problem into a typical large deviation problem: *for an i.i.d. sequence* $\{Y_i(\mathbf{a})\}_{i=1}^n$ *of random variables with* $E[Y_i(\mathbf{a})] = 0$*, estimating the range of* $\sum_{i=1}^n Y_i(\mathbf{a})$.

3 Solution of the problem

With a view toward application, we can ask a more general question: what is the condition number of the following random matrix?

$$\begin{bmatrix} p_0(X_1) & p_1(X_1) & \cdots & p_m(X_1) \\ p_0(X_2) & p_1(X_2) & \cdots & p_m(X_2) \\ \cdots & \cdots & \cdots & \cdots \\ p_0(X_n) & p_1(X_n) & \cdots & p_m(X_n) \end{bmatrix}$$

Here $(X_i)_{i=1}^n$ is an i.i.d. sequence of random variables with values in $\mathbb{R}^d$, and $\{p_0(x), p_1(x), \cdots, p_m(x)\}$ is a linearly independent subset of some functional space over $\mathbb{R}^d$.

Bass and Gröchenig [1] proved the following beautiful theorem. Let $\mathcal{S}$ be a compact subset of the d-dimensional Euclidean space $\mathbb{R}^d$ and let ν be a probability measure on $\mathcal{S}$ with $\mathrm{supp}(\nu) \subset \mathcal{S}$. Let $\mathcal{B}$ be a finite-dimensional subspace of $L^2(\mathcal{S}, \nu)$ with a basis $\{e_k : k = 1, \cdots, m\}$ of continuous functions.

The random sampling problem is to interpolate a given data set $\{(x_i, y_i) : i = 1, \cdots, n\}$ such that there exists an element $p(\cdot; \mathbf{a}) = \sum_{k=1}^m a_k e_k \in \mathcal{B}$ satisfying $p(x_i; \mathbf{a}) = y_i$ $(i = 1, \cdots, n)$. Here the sequence $(x_i)_{i=1}^n$ are i.i.d. samples of the distribution ν.

Theorem 3.1 (**The Random Sampling Theorem**). *The random sampling problem has a unique solution and can be solved using numerically stable algorithms. More precisely, let $\mathcal{U}$ be the $n \times m$ random matrix*

$$\begin{bmatrix} e_1(X_1) & e_2(X_1) & \cdots & e_m(X_1) \\ e_1(X_2) & e_2(X_2) & \cdots & e_m(X_2) \\ \vdots & \vdots & \vdots & \vdots \\ e_1(X_n) & e_2(X_n) & \cdots & e_m(X_n) \end{bmatrix}$$

where $(X_i)_{i=1}^n$ is an i.i.d. sequence of random variables with distribution ν. Then $\mathcal{U}$ is a.s. of full-rank, provided that $n \geq m$ and ν is absolutely continuous w.r.t. Lebesgue measure,

Moreover, there exists constants $A, B > 0$ depending on $\mathcal{S}$, ν and m, such that for all $\rho \in (0, 1)$, the sampling inequality

$$(1 - \rho)n\|p\|_{2,\nu}^2 \leq \sum_{i=1}^n |p(X_i)|^2 \leq (1 + \rho)n\|p\|_{2,\nu}^2, \ \forall p \in \mathcal{B}$$

($\|p\|_{2,\nu}$ denotes the $L^2(\mathcal{S}, \nu)$-norm of p) holds with probability at least

$$1 - Ae^{-Bn\frac{\rho^2}{1+\rho}}.$$

With the same probability estimate, $\kappa(\mathcal{U}) \leq \frac{\beta}{\alpha}\sqrt{\frac{1+\rho}{1-\rho}}$, where α and β satisfying $\alpha\|\mathbf{a}\|_2 \leq \|p(\cdot; \mathbf{a})\|_{L^2(\mathcal{S},\nu)} \leq \beta\|\mathbf{a}\|_2$, $\forall p \in \mathcal{B}$.

Main idea of the proof: The finite dimensional function space $\mathcal{B}$ is a compact subset of $L^2(\mathcal{S}, \nu)$. So for any $\delta > 0$, there is a finite δ-net in $\mathcal{B}$. That is, there exists $\mathcal{A}(\delta) = (p_j(\delta))_{j=1}^m \subset \mathcal{B}$, such that $\forall p \in \mathcal{B}$, $\mathrm{dist}(\mathcal{A}(\delta), p) \leq \delta$. So for any $p = \sum_{k=1}^m a_k e_k \in \mathcal{B}$, we can use elements in those δ-nets to approximate p. Then estimating $P\left(|\sum_{i=1}^n Y_i(p)| > \lambda\right)$ for arbitrary p is reduced to estimating similar probabilities for elements in $\mathcal{B}$. Then we apply probability inequalities.

4 Numerical experiments

We consider the following random matrix

$$A_n = \begin{bmatrix} p_0(X_1) & p_1(X_1) & p_2(X_1) \\ p_0(X_2) & p_1(X_2) & p_2(X_2) \\ \cdots & \cdots & \cdots \\ p_0(X_n) & p_1(X_n) & p_2(X_n) \end{bmatrix}$$

where $(X_i)_{i=1}^n$ is an i.i.d. sequence of random variables following standard normal distribution and p_0, p_1, p_2 are the first three of the Hermite polynomials over $(-\infty, \infty)$:

$$\begin{cases} p_0(x) = 1 \\ p_1(x) = x \\ p_2(x) = \frac{1}{\sqrt{2}}(x^2 - 1) \end{cases}$$

The first experiment lets n take values in $[50, 100]$ and plot $\ln P(\kappa(A_n) > z)$ against n for given z. The result is recorded in Figure 1, which shows

$$\ln P(\kappa(A_n) > z) \approx -nI(z)$$

for some rate function $I(z)$. Figure 2 plots the rate function $I(z)$ for $n = 100$, which is fit by polynomial $-0.085846z^3 + 0.40819z^2 - 0.58275z + 0.26239$ in Figure 3. The mean of the condition number of A_{100} is 1.3527.

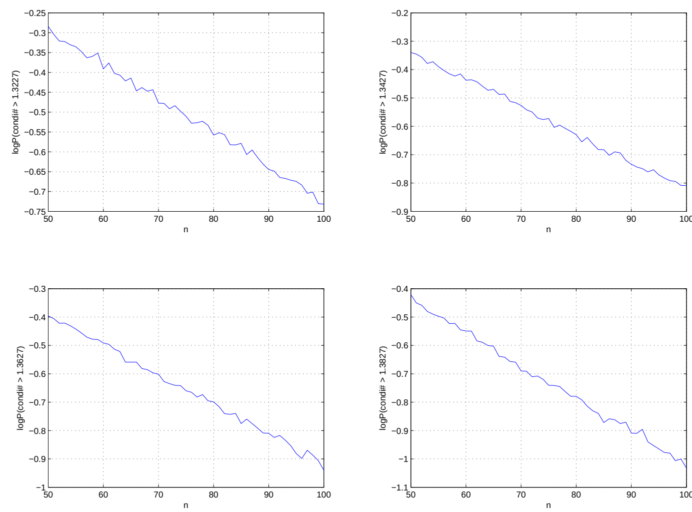

Figure 1: log tail of condition number $\kappa(A_n)$ vs number of samples n

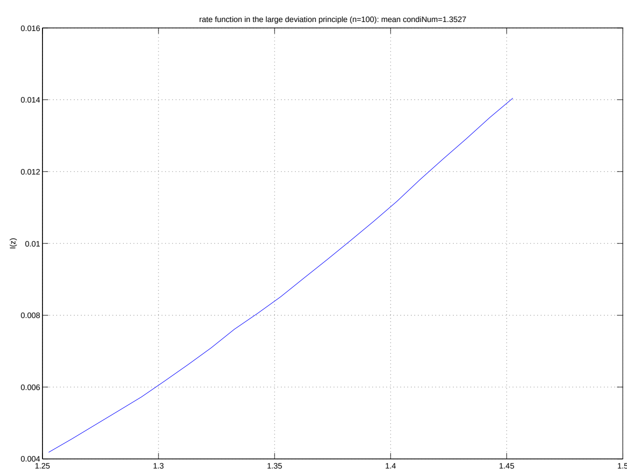

Figure 2: rate function for $n = 100$

The second experiment lets n take values in $[100, 1000]$ and plot $\ln P(\kappa(A_n) > z)$ against n for given z. The result is recorded in Figure 4, which shows

$$\ln P(\kappa(A_n) > z) \approx -nI(z)$$

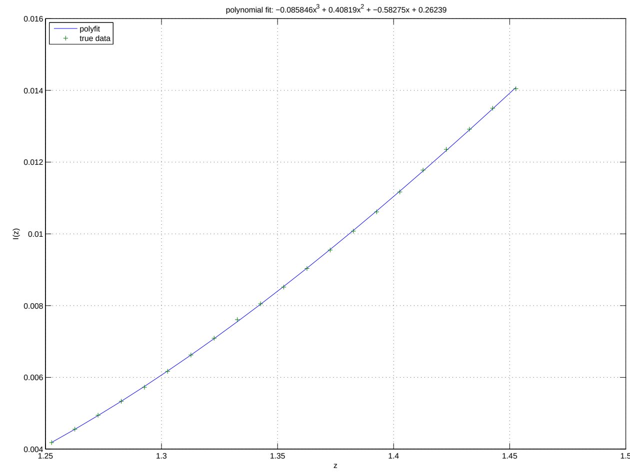

Figure 3: polynomial fit of the rate function

for some rate function $I(z)$. Figure 5 plots the rate function $I(z)$ for $n = 1000$, which is fit by polynomial $-0.071774z^3 + 0.33029z^2 - 0.44765z + 0.18913$ in Figure 6. The mean of the condition number of A_{1000} is 1.0983.

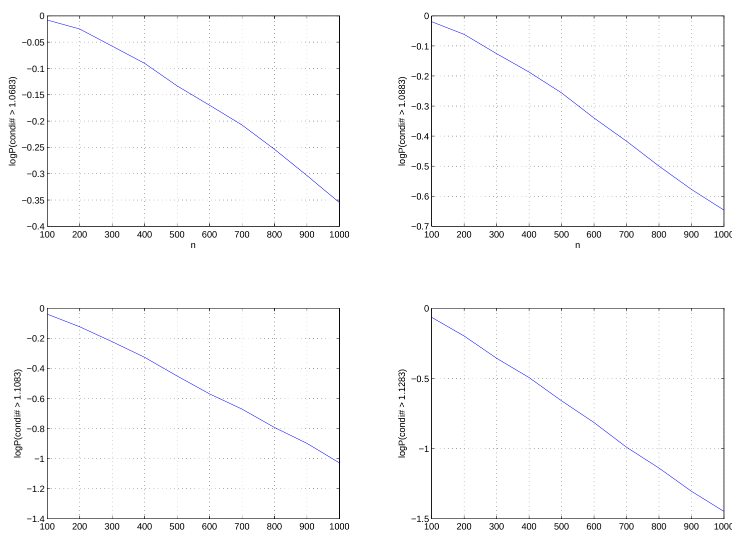

Figure 4: log tail condition number $\kappa(A_n)$ vs number of samples n

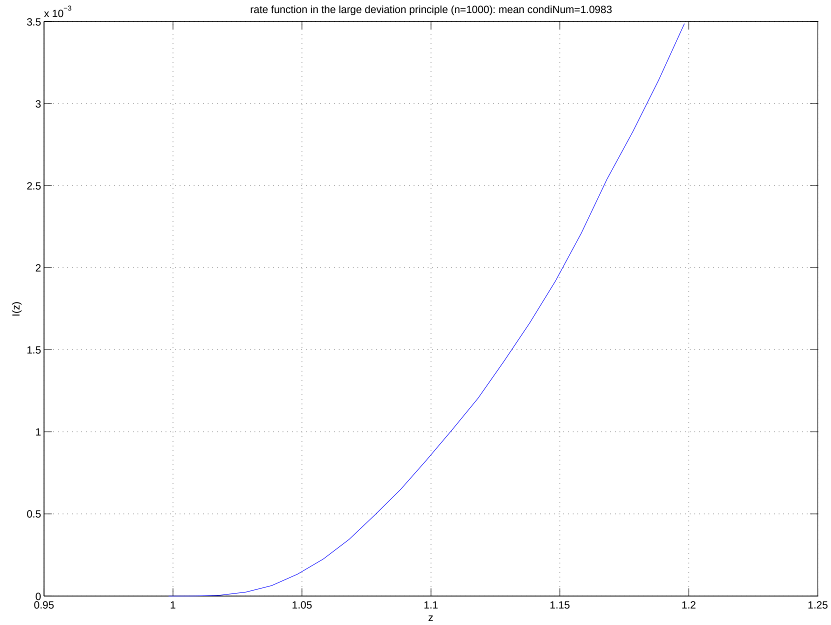

Figure 5: rate function for $n = 1000$

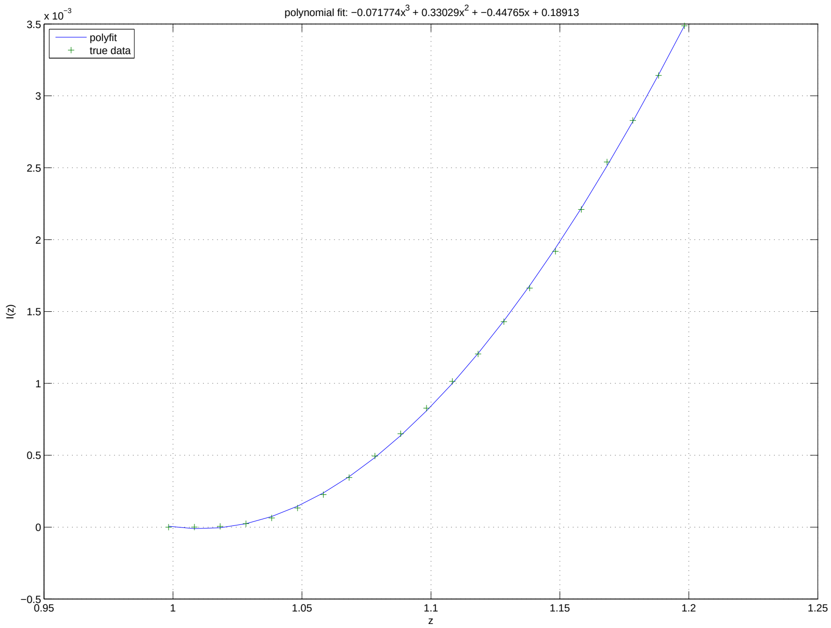

Figure 6: polynomial fit of the rate function

A Matlab code for plotting

```
function con_large_dev

% This code is to verify the condition number of a rectangular random
% Vandermonde-like matrix satisfies the large devation principle.
%
% A typical row of the matrixl will be of the form [H0(x), H1(x), H2(x)],
% where H0, H1, H2 are the first three Hermite functions, and x will be
% i.i.d. samples of a standard Gaussian distribution.
%
% Yan Zeng, 05/20/2011.
%
% Reference
%
% Richard F. Bass and Karlheinz Grochenig. Random sampling of multivariate
% trigonometric polynomials.

% part II. For given n, find the rate function via the formula
% I(z) =  -log[P(condi#>z)] / n.
figure(2);
n = 1000;
numMC = 10000;
condiNum = zeros(numMC, 1);
for i = 1:numMC
    x = randn(n,1);
    condiNum(i) = condiVandermonde_Hermite(x);
end
m = mean(condiNum);
z = (-0.1+m):0.01:(0.1+m);
Iz = zeros(1, length(z));
for j = 1:length(z)
    Iz(j) = log(sum(condiNum>z(j))/numMC);
    Iz(j) = -Iz(j)/n;
end

plot(z, Iz);
xlabel('z');
ylabel('I(z)');
title(['rate function in the large deviation principle (n=', ...
    num2str(n),'): mean condiNum=', num2str(m)]);
grid on;

% polynomial fit of the rate function
figure(3);
[p, ErrorEst] = polyfit(z, Iz, 3);
rate_fit = polyval(p,z,ErrorEst);
plot(z, rate_fit,'-',z,Iz,'+');
legend('polyfit', 'true data', 2);
xlabel('z');
ylabel('I(z)');
title(['polynomial fit: ', num2str(p(1)), 'x^3 + ', num2str(p(2)),....
    'x^2 + ', num2str(p(3)),'x + ' num2str(p(4))]);
```

```
grid on;

% part I. For given x, verify log[P(condi#>z)] = -n*I(z), where n is the
% number of i.i.d. samples, which is also the number of rows in the random
% matrix.
figure(1);
z = (-0.03+m):0.02:(0.03+m);
plotNum = ceil(length(z)/2);
for j = 1:length(z)
    n = 100:100:1000;
    logProb = zeros(1, length(n));
    for k = 1:length(n)
        % Monte Carlo simulation to get probability
        numMC = 10000;
        logProb(k) = calc_logProb(n(k), numMC, z(j));
    end
    subplot(plotNum,2,j);
    plot(n, logProb);
    xlabel('n');
    ylabel(['logP(condi# > ',num2str(z(j)),')']);
    grid on;
end

end %con_large_dev

%%%%%%%%%%% nested functions %%%%%%%%%%%%%
function condiNum = condiVandermonde_Hermite(x)

n = length(x);
A = [ones(n, 1), x, 1/sqrt(2)*(x.*x-1)];
[U, S, V] = svd(A);
condiNum = S(1,1)/S(3,3);

end %condiVandermonde

function p = calc_logProb(n, numMC, z)

condiNum = zeros(numMC, 1);
for i = 1:numMC
    x = randn(n, 1);
    condiNum(i) = condiVandermonde_Hermite(x);
end

p = log(sum(condiNum>z)/numMC);
end %logProb
```

References

[1] Richard F. Bass and Karlheinz Gröchenig. "Random sampling of multivariate trigonometric polynomials". *SIAM J. Math. Anal.*, 2004. 1, 3

[2] Francis A. Longstaff and Eduardo S. Schwartz. "Valuing American options by simulation: a simple least-squares approach". *The Review of Financial Studies*, vol 14, no. 1, pp. 113-147, 2001. 1

[3] Walter Gautschi. "*How (unstable) are Vandermonde systems?*, in Asymptotic and Computational Analysis", R. Wong, ed., vol. 124 of Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, Inc., New York and Basel, 1990, pp. 193-210. 1