Applied Mathematics and Computation 270 (2015) 802–807

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Precise and fast computation of generalized Fermi–Dirac integral by parameter polynomial approximation

Toshio Fukushima*

National Astronomical Observatory of Japan / SOKENDAI (Graduate University of Advanced Studies), 2-21-1, Ohsawa, Mitaka, Tokyo 181-8588, Japan

ARTICLE INFO

Keywords:
Fermi–Dirac integral
Generalized Fermi–Dirac integral
Minimax polynomial approximation
Sommerfeld expansion

ABSTRACT

The generalized Fermi–Dirac integral, $F_k(\eta, \beta)$, is approximated by a group of polynomials of β as $F_k(\eta, \beta) \approx \sum_{j=0}^{J} g_j \beta^j F_{k+j}(\eta)$ where $J = 1(1)10$. Here $F_k(\eta)$ is the Fermi-Dirac integral of order k while g_j are the numerical coefficients of the single and double precision minimax polynomial approximations of the generalization factor as $\sqrt{1 + x/2} \approx \sum_{j=0}^{J} g_j x^j$. If β is not so large, an appropriate combination of these approximations computes $F_k(\eta, \beta)$ precisely when η is too small to apply the optimally-truncated Sommerfeld expansion (Fukushima, 2014 [15]). For example, a degree 8 single precision polynomial approximation guarantees the 24 bit accuracy of $F_k(\eta, \beta)$ of the orders, $k = -1/2(1)5/2$, when $-\infty < \eta \le 8.92$ and $\beta \le 0.2113$. Also, a degree 7 double precision polynomial approximation assures the 15 digit accuracy of $F_k(\eta, \beta)$ of the same orders when $-\infty < \eta \le 29.33$ and $0 \le \beta \le 3.999 \times 10^{-3}$. Thanks to the piecewise minimax rational approximations of $F_k(\eta)$ (Fukushima, 2015 [18]), the averaged CPU time of the new method is roughly the same as that of single evaluation of the integrand of $F_k(\eta, \beta)$. Since most of $F_k(\eta)$ are commonly used in the approximation of $F_k(\eta, \beta)$ of multiple contiguous orders, the simultaneous computation of $F_k(\eta, \beta)$ of these orders is further accelerated by the factor 2–4. As a result, the new method runs 70–450 times faster than the direct numerical integration in practical applications requiring $F_k(\eta, \beta)$.

1. Introduction

The generalized Fermi–Dirac integral is an important special function in astrophysics [1, Section 24.4] and in solid state physics [2]. The integral is written in the form of an integral transform [3, Section 6.10] as

$$F_k(\eta, \beta) \equiv \int_0^\infty \frac{t^k \sqrt{1 + (\beta/2)t}\, dt}{\exp(t - \eta) + 1}. \quad (k > -1;\ -\infty < \eta < \infty;\ \beta \ge 0) \tag{1}$$

It is a generalization of $F_k(\eta)$, the Fermi–Dirac integral [2] widely used in quantum statistics, in the sense $F_k(\eta) = F_k(\eta, 0)$. There are a few other expressions of the generalized Fermi–Dirac integrals [4,5]. However, the above standard form will be treated throughout this article.

The precise and fast computation of $F_k(\eta, \beta)$ is quite difficult [3, Section 6.10]. This issue has been discussed since the early days when a Nobel laureate, Chandrasekhar, determined the limit mass of white dwarf by employing the integrals [6]. Indeed, a single chapter is fully devoted to its computation in a modern text on the stellar structure [1, Chapter 24].

* Corresponding author. Tel.: +81422343613.
E-mail address: Toshio.Fukushima@nao.ac.jp

http://dx.doi.org/10.1016/j.amc.2015.08.094
0096-3003/

Despite that many methods have been proposed to compute the integral [5,7–12], the current standard procedure [3, Section 6.10] is the numerical quadrature by the double exponential rule [13] with a splitting of the integration interval at $t = \eta$. The method is sufficiently accurate but slow in general. In the astrophysical applications, frequently required is the simultaneous computation of $F_k(\eta, \beta)$ of three contiguous half integer orders, $k = 1/2,\ 3/2$, and $5/2$. As a result, the standard method requires 150–1800 evaluations of the integrand for each combination of η and β [3, Section 6.10].

This is quite time-consuming if considering their frequent utilization. For example, the evaluation of the integrals must be conducted at all locations inside the stars at every step of their evolution tracking. Thus, the evolution computation of a three-dimensional star with even a small number of spatial grids as 128 for just 1000 time steps would require the number of the simultaneous evaluation of the three integrals as many as $128^3 \times 1000 \approx 2.1 \times 10^9$. Even if each computation costs only 50 μs, the total CPU time would be of the order of day.

Recently, an extension of the numerical integration method of [14, Section 3] is developed to compute a general integral of the Fermi–Dirac distribution including $F_k(\eta, \beta)$ [15]. This method significantly reduces the number of integrand evaluations when $\eta > 4$ by the factor 1.2–120 [15, Fig. 3]. Nevertheless, the extended method is not effective for negative and small positive values of η. On the other hand, when η is sufficiently large, applicable is the Sommerfeld expansion [16] to compute a general integral of the Fermi–Dirac distribution. Although the expansion is of asymptotic nature, its appropriate truncations compute $F_k(\eta, \beta)$ with the single and double precision accuracies, respectively [17]. However, still challenging is the precise and fast computation of $F_k(\eta, \beta)$ for not so large values of η.

In order to overcome this situation, by using newly-developed procedures to compute $F_k(\eta)$ [18], a new method is developed to compute $F_k(\eta, \beta)$ in this article. Below, Section 2 explains the new method of computation, and Section 3 shows the results of numerical experiments.

2. Method

2.1. Series expansion of $F_k(\eta, \beta)$ with respect to β

Consider the computation of $F_k(\eta, \beta)$ of the half integer orders, $-1/2(1)5/2$, when η is less than η_S, the threshold values of the truncated Sommerfeld expansion [17], which ranges 8.9–13.4 and 29.3–38.8 in the single and the double precision environments, respectively. In many practical applications, the parameter β is relatively small. In the astrophysical cases, for example, the maximum value of β of electrons in ordinary stars like the Sun is of the order of 10^{-3} [7]. Meanwhile, that of protons as well as neutrons is less than 0.1 for all cases. On the other hand, also small are β of typical semiconductors including that of InSb being around 0.3 [8].

Therefore, examined is the series expansion of $F_k(\eta, \beta)$ with respect to β. The key function here is the generalization factor, $g(x) \equiv \sqrt{1 + x/2}$, where $x \equiv \beta t$. If it is well approximated as a polynomial of x as $g(x) \approx \sum_{j=0}^{J} g_j x^j$, then an approximation of $F_k(\eta, \beta)$ is obtained by the termwise integration as

$$F_k(\eta, \beta) \approx F_k^{(J)}(\eta, \beta) \equiv \sum_{j=0}^{J} g_j \beta^j F_{k+j}(\eta). \tag{2}$$

Notice that the coefficients g_j are independent on the order k, and therefore, they can be commonly used for the approximation of $F_k(\eta, \beta)$ of different orders in general.

2.2. Approximation of generalization factor

Consider an approximation of the generalization factor, $g(x)$. Since the effectiveness of the Maclaurin series expansion [21, Formula 15.4.6] is limited by an algebraic singularity of $g(x)$ at $x = -2$, developed are the minimax polynomial approximations. Assume that both of J, the degree of the approximation polynomial, and X, the length of the approximation interval, are specified. Then, a set of the polynomial coefficients, g_j for $j = 0, 1, \ldots J$, are determined so as to minimize the relative approximation error in the interval $0 \le x \le X$. In practice, the process is reversed. Namely, when the required error tolerance is specified, X is maximized under the condition the maximum relative error of approximation does not exceed the tolerance. The reversed procedure implemented by Mathematica Version 10 [19] determined the coefficients as listed in Tables 1–3.

2.3. Maximum parameter

Now that the minimax polynomial approximation of $g(x)$ is established, easily constructed are the polynomial approximations of $F_k(\eta, \beta)$ by using them. Since the interval of η is as limited as $-\infty < \eta \le \eta_S$, the maximum relative error for the given value of β is measured by setting $\eta = \eta_S$. This must be compared with δ, the given relative error tolerance. Extending the process, actually determined is β_S, the maximum value of β when $\eta \le \eta_S$ such that the relative error does not exceed δ. The relative errors of $F_k(\eta, \beta)$ are measured by numerically comparing the approximate value with the reference values, which are prepared by the quadruple precision numerical integration by means of a quadruple precision extension of Ooura's `intde`, an adaptive numerical quadrature program in the double precision environment based on the double exponential rule [20]. Notice that, only the integrals of negative arguments are required in some applications such as the evaluation of the physical quantities of the