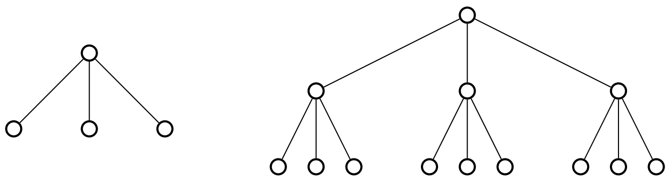
A *rooted tree* is a tree together with a designated vertex r called the *root*. A *complete k-ary tree* of depth d is defined recursively as follows:

- A complete k-ary tree of depth 0 is a single vertex.
- For $d \geq 0$, a complete k-ary tree of depth $d+1$ is obtained by taking k complete k-ary trees $T_1, \dots, T_k$ of depth d, a new root vertex r, and adding edges from r to the roots of $T_1, \dots T_k$.

Here are diagrams of the complete ternary (3-ary) trees of depth 1 and 2:



How many vertices $N(d)$ does a complete k-ary tree of depth d have? When $d \geq 1$, there is one vertex for each of the k subtrees of depth $d-1$, plus the root vertex. This gives the formula

$$N(d) = k \cdot N(d-1) + 1$$

for $d \geq 1$, with the "base case" $N(0) = 1$. Plugging in small values of d, this gives

$$\begin{aligned}
N(1) &= k \cdot N(0) + 1 = k + 1 \\
N(2) &= k \cdot N(1) + 1 = k(k+1) + 1 = k^2 + k + 1 \\
N(3) &= k \cdot N(2) + 1 = k(k^2 + k + 1) + 1 = k^3 + k^2 + k + 1
\end{aligned}$$

and, in general,

$$N(d) = k^d + k^{d-1} + \cdots + 1 \qquad \text{for every } d \geq 0.$$

You can prove the correctness of this formula by induction on d, but we won't bother. Today we are more interested in "understanding" the value of $N(d)$.

1 Evaluating sums

Geometric sums In general, we can evaluate a sum of the form

$$S = x^d + x^{d-1} + \cdots + 1$$

for every real number x and positive integer d like this: If we multiply both sides by x, we obtain

$$xS = x^{d+1} + x^d + \cdots + x$$

If we now subtract the first expression from the second one, almost all the right hand sides terms cancel out:

$$xS - S = x^{d+1} - 1$$

which simplifies to $(x-1)S = x^{d+1} - 1$. When $x \neq 1$, we can do a division and obtain the formula

$$x^d + x^{d-1} + \cdots + 1 = \frac{x^{d+1} - 1}{x - 1} \qquad \text{for every real number } x \neq 1.$$

A sum of this form is called a *geometric sum.*

So the number of vertices in a complete k-ary tree of depth d is $(k^{d+1} - 1)/(k-1)$. In particular, for a complete ternary tree, this number is $(3^{d+1} - 1)/2$. A complete binary (2-ary) tree of depth d has $2^{d+1} - 1$ vertices.

Polynomial sums In Lecture 3 we proved that

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

for every integer $n \geq 0$. How did I come up with the expression on the right? Instead of going back to something we already know, let's work a new one together:

$$\text{What is } 1^2 + 2^2 + \cdots + n^2?$$

We have to do a bit of guesswork here. Since the sum $1+2+\cdots+n$ turned out to equal a quadratic polynomial in n, perhaps $1^2 + 2^2 + \cdots + n^2$ might equal a cubic polynomial? Let's make a guess: For all n, there exist real numbers a, b, c, d such that

$$1^2 + 2^2 + \cdots + n^2 = an^3 + bn^2 + cn + d.$$

Suppose our guess was correct. Then what are the numbers a, b, c, d? We can get an idea by evaluating both sides for different values of n:

$$\begin{aligned} 0 &= d && \text{for } n = 0 \\ 1 &= a + b + c + d && \text{for } n = 1 \\ 5 &= 8a + 4b + 2c + d && \text{for } n = 2 \\ 14 &= 27a + 9b + 3c + d && \text{for } n = 3. \end{aligned}$$

I solved this system of equations on the computer and obtained $a = 1/3$, $b = 1/2$, $c = 1/6$, $d = 0$. This suggests the formula

$$1^2 + 2^2 + \cdots + n^2 = \tfrac{1}{3}n^3 + \tfrac{1}{2}n^2 + \tfrac{1}{6}n$$

for all integers $n \geq 0$. Let us see if we can prove its correctness by induction on n.

We already worked out the base case $n = 0$, so let us do the inductive step. Fix $n \geq 0$ and assume that the equality holds for n. Then

$$1^2 + 2^2 + \cdots + (n+1)^2 = \left(\tfrac{1}{3}n^3 + \tfrac{1}{2}n^2 + \tfrac{1}{6}n\right) + (n+1)^2 = \tfrac{1}{3}n^3 + \tfrac{3}{2}n^2 + \tfrac{13}{6}n + 1.$$

This indeed equals $\frac{1}{3}(n+1)^3 + \frac{1}{2}(n+1)^2 + \frac{1}{6}(n+1)$. So we have discovered and proved a new theorem:

Theorem 1. *For every integer $n \geq 0$, $1^2 + 2^2 + \cdots + n^2 = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$.*

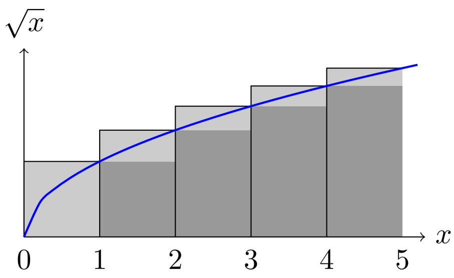
2 Approximating sums

Exact "closed-form" expressions for sums are rather exceptional. Fortunately, we can often obtain very good approximations. One powerful method for approximating sums is the integral method from calculus: It works by comparing the sum with a related integral.

As an example, let us look at the expression:

$$S(n) = \sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}.$$

For example, $S(5)$ equals the area covered by the shaded rectangles (both light and dark shades) in the following plot:



The area under the rectangles can be lower bounded by the area (i.e., the integral) of the curve $f(x) = \sqrt{x}$ from $x = 0$ to $x = 5$:

$$S(5) \geq \int_0^5 \sqrt{x}\, dx.$$

If we remove the area L covered by the lightly shaded rectangles, the darker shaded area is now dominated by the curve $f(x) = \sqrt{x}$ and so

$$S(5) - L \leq \int_0^5 \sqrt{x}\, dx.$$

The area under L is exactly $\sqrt{5}$: If we stack all of the lightly shaded rectangles on top of one another, we obtain a column of width 1 and height $\sqrt{5}$. Therefore

$$\int_0^5 \sqrt{x}\, dx \leq S(5) \leq \int_0^5 \sqrt{x}\, dx + \sqrt{5}.$$

By the same reasoning, for every integer $n \geq 1$, we have the inequalities

$$\int_0^n \sqrt{x}\, dx \leq S(n) \leq \int_0^n \sqrt{x}\, dx + \sqrt{n}.$$

We can now use rules from calculus to evaluate the integrals: Recalling that $x^{1/2} = \frac{d}{dx}\frac{2}{3}x^{3/2}$, it follows from the fundamental theorem of calculus that

$$\frac{2}{3}n^{3/2} \leq S(n) \leq \frac{2}{3}n^{3/2} + \sqrt{n}.$$

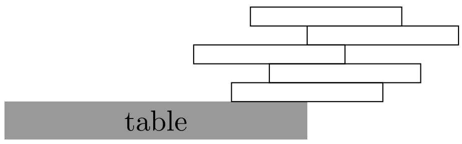
To get a feel for these inequalities, let us plug in a few values of n. (I calculated $S(n)$ by evaluating the sum on the computer.)

n	$\frac{2}{3}n^{3/2}$	$S(n)$	$\frac{2}{3}n^{3/2} + \sqrt{n}$
10	21.082	22.468	24.244
100	666.67	671.46	676.67
1,000	21,081.9	21,097.5	21,113.5
10,000	666,666	666,716	666,766

As n becomes large, the accuracy of these approximations looks quite good.

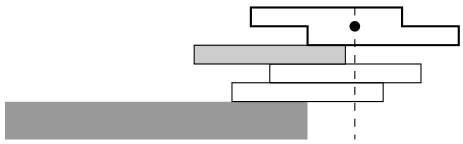
3 Overhang

You have n identical rectangular blocks and you stack them on top of one another at the edge of a table like this:



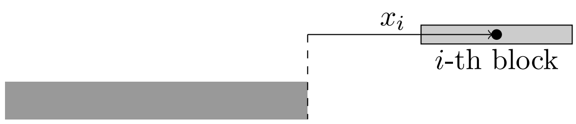
Is this configuration stable, or will it topple over?

In general, a configuration of n blocks is *stable* if for every i between 1 and n, the center of mass of the top i blocks sits over the $(i + 1)$st block, where we think of the table as the $(n + 1)$st block in the stack. For example, the top stack is not stable because the center of mass of the top two blocks does not sit over the third block:



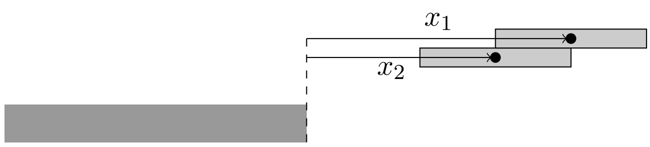
We want to stack our n blocks so that the rightmost block hangs as far over the edge of the table as possible. What should we do? One reasonable strategy is to try to push the top blocks as far as possible away from the table as long as they do not topple over.

We will assume each block has length 2 units and we will use x_i to denote the offset of the center of the i-th block (counting from the top) from the edge of the table:



The offset of a block can be positive, zero, or negative, depending on the position of its center of mass.

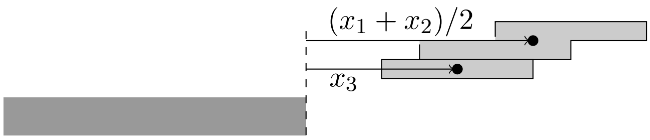
For the top block not to topple, its center of mass must sit over the second block. To move it as far away from the edge of the table as possible, we move its center exactly one unit to the right of the center of the second block:



This forces the offsets x_1 and x_2 to satisfy the equation

$$x_1 = x_2 + 1. \tag{1}$$

Now we move onto the third block. The center of mass of the first two blocks is at offset $(x_1 + x_2)/2$ from the edge of the table. To push this as far to the right as possible without toppling over the third block



we must set

$$\frac{x_1 + x_2}{2} = x_3 + 1. \tag{2}$$

Continuing our reasoning in this way, for every i between 1 and n, the offset of the center of mass of the top i blocks is $(x_1 + \cdots + x_i)/i$. To push this as far to the right without toppling over the $(i+1)$st block, we must set

$$\frac{x_1 + x_2 + \cdots + x_i}{i} = x_{i+1} + 1 \quad \text{for all } 1 \leq i \leq n. \tag{3}$$

Finally, when $i = n + 1$, we have reached the table whose offset is zero. Since we are thinking of the table as the $(n+1)$st block, its centre of mass is one unit left to its edge:

$$x_{n+1} = -1. \tag{4}$$

The overhang of the set of blocks is $x_1 + 1$. To figure out what this number is, we need to solve for x_1 in the system of equations (3-4). Let us develop some intuition first. Equation (1) tells us that $x_2 = x_1 - 1$. Plugging in this formula for x_2 into (2), we get that

$$x_3 = x_1 - \frac{1}{2} - 1 = x_1 - \frac{3}{2}.$$

Let's do one more step. Equation (3) tells us that $(x_1 + x_2 + x_3)/3 = x_4 + 1$. Plugging in our formulas for x_2 and x_3 in terms of x_1 we get that

$$\frac{x_1 + (x_1 - 1) + (x_1 - \frac{3}{2})}{3} = x_4 + 1$$

from where

$$x_4 = x_1 - \frac{1 + (\frac{3}{2})}{3} - 1 = x_1 - \frac{11}{6}.$$

We get $x_2 = x_1 - 1$, $x_3 = x_1 - \frac{3}{2}$, $x_4 = x_1 - \frac{11}{6}$. There doesn't seem to be any pattern. But if you look up the sequence $1, 3/2, 11/6$ in the *Online encyclopedia of integer sequences*, you find out that there is a nice representation of these numbers:

$$1 = 1 \qquad \frac{3}{2} = 1 + \frac{1}{2} \qquad \frac{11}{6} = 1 + \frac{1}{2} + \frac{1}{3}.$$

We can now make a guess about what the solution looks like and try to prove that our guess is correct:

Theorem 2. *The assignment*

$$x_{i+1} = x_1 - 1 - \frac{1}{2} - \cdots - \frac{1}{i}$$

for i ranging from 1 to n satisfies all the equations (3).

Proof. Take any i between 1 and n. Under this assignment, the left hand side of the i-th equation equals

$$\frac{1}{i}\Big(x_1 + (x_1 - 1) + \cdots + \Big(x_1 - 1 - \frac{1}{2} - \cdots - \frac{1}{i-1}\Big)\Big) = \frac{1}{i}\Big(i \cdot x_1 - (i-1) \cdot 1 - (i-2) \cdot \frac{1}{2} - \cdots - 1 \cdot \frac{1}{i-1}\Big)$$

because there are i instances of x_1, $i-1$ instances of -1, $i-2$ instances of $1/2$, and so on. If we pull out all the factors of i, we obtain that this equals

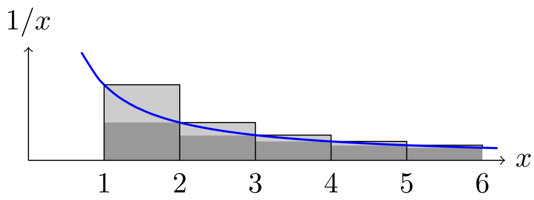
$$\frac{1}{i}\Big(i \cdot x_1 - i \cdot 1 - \cdots - i \cdot \frac{1}{i-1}\Big) + \frac{1}{i}\Big(1 \cdot 1 + 2 \cdot \frac{1}{2} + \cdots + (i-1) \cdot \frac{1}{i-1}\Big).$$

The first term equals $x_{i+1} + 1/i$. The second term equals, after cancellation, equals $(i-1)/i$. So their sum equals $x_{i+1} + 1$, which shows that the i-th equation is satisfied. □

Since $x_{n+1} = -1$, Theorem 2 tells us that $x_1 = 1/2 + \cdots + 1/n$, so the overhang for this configuration of blocks is

$$1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

This number is called the *n-th harmonic number* $H(n)$. There is no closed form formula for it but we can obtain an excellent approximation using the integral method. To do this, we compare $H(n)$ with the integral of the function $1/x$:



By similar reasoning as before, the sum $H(n) = 1 + 1/2 + \ldots 1/n$ is given by the area of the first n shaded rectangles. This area is larger than the integral of $1/x$ from 1 to $n+1$:

$$H(n) \geq \int_1^{n+1} \frac{1}{x}\, dx.$$

On the other hand, if we subtract from $H(n)$ the area of the lightly shaded rectangles, then the integral becomes larger. This area equals $1 - 1/(n+1)$:

$$H(n) - 1 + \frac{1}{n+1} \leq \int_1^{n+1} \frac{1}{x}\, dx.$$

Combining this two inequalities gives the approximation

$$\int_1^{n+1} \frac{1}{x}\, dx \leq 1 + \frac{1}{2} + \cdots + \frac{1}{n} \leq \int_1^{n+1} \frac{1}{x}\, dx + 1 - \frac{1}{n+1}.$$

The antiderivative of $1/x$ is $\ln x$. By the fundamental theorem of calculus it follows that

$$\ln(n+1) \leq H(n) \leq \ln(n+1) + 1 - \frac{1}{n+1}.$$

The left hand side of this inequality tells us that our method of stacking blocks achieves overhang at least $1 + \ln(n+1)$. The logarithm function is unbounded; given enough blocks, we can grow our stack to reach all the way to New York!

4 Order of growth

In engineering we are often interested in the *asymptotic* behaviour of measures as our problem size becomes large. For example, if we have a computer program for routing packets through a switching network, we might not really care what happens for 2 or 3 packets, but would want to have a good understanding about how fast the program is when we have hundreds or thousands of packets, or even more in the era of "big data". For this purpose, it is useful to have a quick way of roughly figuring out how a function $f(n)$ behaves as n grows large. Usually, we do this by comparing the value of the function f for large n to values of "well-known" functions like n, n^2, $\log n$, 2^n, or e^n.

The big-oh notation is a handy way to say that a given function does not grow too fast.

Definition 3. For two real-valued functions f and g (defined over the positive reals, or over the positive integers), we say f is $O(g)$ (big-oh of g) if there exists a constant $C > 0$ such that for every sufficiently large input x, $f(x) \le C \cdot g(x)$.

For example, $13x^4 + 2x^2 + 10x + 1000$ is $O(x^4)$ because when x is large (specifically, at least 1):

$$13x^4 + 2x^2 + 10x + 1000 \le 13x^4 + 2x^4 + 10x^4 + 1000x^4 = 1025x^4.$$

By the same reasoning, every polynomial is the big-oh of its leading term.

Similarly, $\log(16x^5 + 3x + 11)$ is $O(\log x)$ because when x is large,

$$\log(16x^5 + 3x + 11) \le \log(16x^5 + 3x^5 + 11x^5) \le \log(30x^5) \le \log(x^6) \le 6 \log x.$$

For the harmonic number $H(n)$ we derived the inequality $H(n) \le \log(n+1) + 1 - 1/(n+1)$. When n is large,

$$H(n) \le \log(n+1) + \left(1 - \frac{1}{n+1}\right) \le \log n^2 + \log n = 2\log n + \log n = 3 \log n$$

so $H(n)$ is $O(\log n)$.

The little-oh notation says that asymptotically, one function grows at a significantly slower rate than another one.

Definition 4. For two real-valued functions f and g (defined over the positive reals, or over the positive integers), we say f is $o(g)$ (little-oh of g) if *for every* constant $c > 0$ and every sufficiently large input x, $f(x) \le c \cdot g(x)$.

If f is $o(g)$, then f is also $O(g)$, but not necessarily the other way. For example, $\frac{1}{2}x^5 - x^3$ is $O(x^5)$, but it is not $o(x^5)$ because as x gets large, $(\frac{1}{2}x^5 - x^3)/x^5$ converges to $\frac{1}{2}$; when x is sufficiently large, $\frac{1}{2}x^5 - x^3$ will be greater than, say, $\frac{1}{4}x^5$.

For example, $x^{3/2} + 3x^{1/2}$ is $o(x^2)$ because

$$x^{3/2} + 3x^{1/2} \le 4x^{3/2} \le \frac{4}{x^{1/2}} \cdot x^2$$

and for every constant $c > 0$, as x becomes large enough, $4/x^{1/2}$ is smaller than c, so $x^{3/2}$ is at most cx^2. By similar reasoning, every polynomial $p(x)$ of degree d (even one with fractional degrees) is $o(x^e)$ for every constant $e > d$.

When $0 < B < C$, it is also true that B^x is $o(C^x)$ because we can write $B^x = (B/C)^x \cdot C^x$ and for any constant c, $(B/C)^x$ becomes eventually smaller than c.

Theorem 5. *For all constants $a, b > 0$, $(\log x)^a$ is $o(x^b)$.*

In the statement of this theorem, the base of the logarithm is irrelevant because changing the base from one constant to another only changes the value of the expression $(\log x)^a$ by a constant factor. So we will assume, without loss of generality, that the logarithm is a natural logarithm.

Proof. When $a = 1$, we can calculate $\lim_{x\to\infty}(\ln x)/x^b$ using L'Hôpital's rule from calculus. Both numerator and denominator grow to infinity, but this is not true for their derivatives: $\frac{d}{dx}\ln x = 1/x$, while $\frac{d}{dx}x^b = bx^{b-1}$. The ratio of these two numbers is $1/(bx^b)$, which tends to zero as x grows. Therefore

$$\lim_{x\to\infty}\frac{\ln x}{x^b} = \lim_{x\to\infty}\frac{1}{bx^b} = 0$$

and so $\log x \le cx^b$ for every constant $c > 0$ and sufficiently large x.

Now let $a > 0$ be arbitrary and $c > 0$ be an arbitrary constant. By what we just proved,

$$\ln x \le c^{1/a}x^{b/a}$$

for x sufficiently large, from where

$$(\ln x)^a \le (c^{1/a}x^{b/a})^a \le cx^b$$

for x sufficiently large. □

If we set $x = e^y$ and $B = e^b$, we get the following corollary:

Corollary 6. *For all constants $a > 0$ and $B > 1$, y^a is $o(B^y)$.*

Big-omega and little-omega are the opposites of big-oh and little-oh, respectively: We say f is $\Omega(g)$ if g is $O(f)$ and we say f is $\omega(g)$ if g is $o(f)$. Finally, we say f is $\Theta(g)$ if f is $O(g)$ and f is $\Omega(g)$. For example, $x^5 + 7x^3$ is $\Theta(x^5)$ and $H(n)$ is $\Theta(\log n)$. You'll practice all these extensively in the homework.

All these relations are *transitive*: For example, if f is $O(g)$ and g is $O(h)$ then f is $O(h)$.

It is customary to abuse the equality sign when talking about order of growth. For example, people often write "$f = O(g)$" instead of "f is $O(g)$". Technically, this is incorrect because f and $O(g)$ are objects of different types: f is a single function while $O(g)$ is not. However it is sometimes very convenient. It is okay to use this notation as long as you are aware of what it means. What you should *not* do is write "equations" like

$$1 + 2 + \cdots + n = O(1) + O(1) + \cdots + O(n) = (n-1)\cdot O(1) + O(n) = O(n)$$

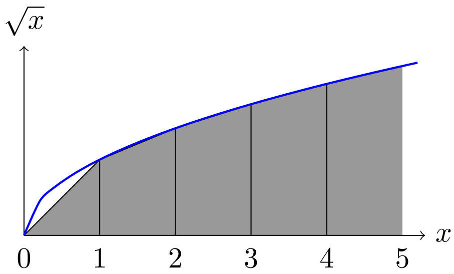
because it is not clear what they mean and they may be incorrect.

5 Better approximations*

The estimate we obtained for $S(n) = 1 + \sqrt{2} + \cdots + \sqrt{n}$ has the undesirable property that the errors $S(n) - \frac{2}{3}n^{3/2}$ and $\frac{2}{3}n^{3/2} + \sqrt{n} - S(n)$ appear to grow to infinity as n becomes larger. You may notice, however, that the average of the upper and lower approximation looks much better:

n	$S(n)$	$\frac{2}{3}n^{3/2}+\frac{1}{2}\sqrt{n}$
10	22.468	24.663
100	671.463	676.666
$1,000$	$21,097.456$	$21,096.662$
$10,000$	$666,716.459$	$666,716.666$

The estimate $\frac{2}{3}n^{3/2}+\frac{1}{2}\sqrt{n}$ can be obtained more methodically as an application of the trapezoidal rule, which estimates the shaded area in this picture by the integral under the curve in the following plot:



Indeed, the shaded area can be calculated by adding the areas under the first n trapezoids to obtain

$$A=\frac{1}{2}(\sqrt{0}+\sqrt{1})+\frac{1}{2}(\sqrt{1}+\sqrt{2})+\cdots+\frac{1}{2}(\sqrt{n-1}+\sqrt{n})=S(n)-\frac{1}{2}\sqrt{n}.$$

This area is upper bounded by the area under the curve $f(x)=\sqrt{x}$ between $x=0$ and $x=n$, so

$$A\leq\int_0^n\sqrt{x}dx=\frac{2}{3}n^{3/2}$$

from where

$$S(n)\leq\frac{2}{3}n^{3/2}+\frac{1}{2}\sqrt{n}.$$

It appears that the approximation error $\frac{2}{3}n^{3/2}+\frac{1}{2}\sqrt{n}-S(n)$ is a small fraction, something like $1/4$, regardless of the value of n. Is this really the case?

To bound the error we need to understand the contribution by a single trapezoid. For convenience we will shift and scale the trapezoid to have x-coordinates -1 and 1. The following clever lemma bounds the error by the second derivative f'' of f.

Lemma 7. *For every function f from the real interval $[-1,1]$ to the real numbers so that $-B\leq f''(x)\leq 0$ for all $-1\leq x\leq 1$,*

$$f(-1)+f(1)\leq\int_{-1}^1 f(t)dt\leq f(-1)+f(1)+\frac{2B}{3}. \tag{5}$$

Proof. The integration by parts formula from calculus says that

$$\int_{-1}^1 f(t)\cdot 1dt+\int_{-1}^1 f'(t)\cdot tdt=f(t)\cdot t\big|_{-1}^1=f(-1)+f(1).$$

Another application of integration by part gives that

$$\int_{-1}^1 f'(t)\cdot tdt+\int_{-1}^1 f''(t)\cdot\tfrac{1}{2}(t^2-1)dt=f'(t)\cdot(t^2-1)\big|_{-1}^1=0$$

because $t^2 - 1$ evaluates to zero at both endpoints. Subtracting the second equation from the first and moving terms around gives the convenient identity

$$\int_{-1}^{1} f(t)dt = f(-1) + f(1) + \int_{-1}^{1} f''(t) \cdot \frac{1}{2}(t^2 - 1)dt.$$

Neither function under the integrand on the right is postivie in the range $[-1, 1]$, so the integral itself must be nonnegative. Therefore $\int_{-1}^{1} f(t)dt \leq f(-1) + f(1)$. On the other hand,

$$\int_{-1}^{1} f''(t) \cdot \tfrac{1}{2}(t^2 - 1)dt \leq \tfrac{1}{2}B \int_{-1}^{1} (1 - t^2)dt = -\tfrac{1}{2}B\big(t - \tfrac{1}{3}t^3\big)\big|_{-1}^{1} = \frac{2B}{3}$$

as desired. □

Applying the change of variables $x = (t + i + 1)/2$ in (5) gives that for every i,

$$\frac{f(i) + f(i+1)}{2} \leq \int_{i}^{i+1} f(x)dx \leq \frac{f(i) + f(i+1)}{2} + \frac{B_i}{3} \tag{6}$$

assuming that $0 \leq f''(x) \leq B_i$ for every x between i and $i + 1$.

If f is the function $f(x) = \sqrt{x}$, its second derivative is $f''(x) = -\frac{1}{4}x^{-3/2}$, so it takes value between $-\frac{1}{4}i^{-3/2}$ and zero in the interval $[i, i + 1]$. Applying (6) gives that

$$\frac{1}{2}(\sqrt{i} + \sqrt{i+1}) \leq \int_{i}^{i+1} \sqrt{x}dx \leq \frac{1}{2}(\sqrt{i} + \sqrt{i+1}) + \frac{1}{12i^{3/2}}.$$

The right hand side is meaningless when $i = 0$, so we start at $i = 1$ instead. Adding up these inequalities for i taking integer values from 1 up to $n - 1$ gives

$$S(n) - \frac{1}{2}(\sqrt{1} + \sqrt{n}) \leq \int_{1}^{n} \sqrt{x}dx \leq S(n) - \frac{1}{2}(\sqrt{1} + \sqrt{n}) + E(n)$$

where

$$E(n) = \frac{1}{12} + \frac{1}{12 \cdot 2^{3/2}} + \cdots + \frac{1}{12 \cdot (n-1)^{3/2}}.$$

Calculating the integral and simplifying gives that

$$\frac{2}{3}n^{3/2} + \frac{1}{2}\sqrt{n} - \frac{1}{6} - E(n) \leq S(n) \leq \frac{2}{3}n^{3/2} + \frac{1}{2}\sqrt{n} - \frac{1}{6}.$$

This estimate appears even better than the previous one as subtracting $1/6 \approx 0.166$ seems to bring the estimate even closer to the true value of the sum $S(n)$.

It remains to analyze the error term $E(n)$. This can be done by another integral bound:

$$E(n) \leq \int_{1}^{n+1} \frac{1}{12x^{3/2}}\, dx = \frac{1}{12} + (-\tfrac{1}{6}x^{-1/2})\big|_{1}^{n+1} \leq \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$$

so the estimate is always within $1/4$ of the true value.

In fact, we can make the approximation error arbitrarily small by starting the summation at a later point. For instance, the same method tells us that

$$\frac{2}{3}n^{3/2} + \frac{1}{2}\sqrt{n} - 16.5 \leq \sqrt{9} + \sqrt{10} + \cdots + \sqrt{n} \leq \frac{2}{3}n^{3/2} + \frac{1}{2}\sqrt{n} - 16.5 + E(9, n)$$

where

$$E(9,n) = \frac{1}{12 \cdot 9^{3/2}} + \frac{1}{12 \cdot 10^{3/2}} + \cdots + \frac{1}{12 \cdot n^{3/2}} \le \int_8^\infty \frac{1}{12x^{3/2}} dx = \frac{1}{12 \cdot 8^{2/3}} \le 0.0834.$$

Therefore the formula

$$\frac{2}{3}n^{3/2} + \frac{1}{2}\sqrt{n} - 16.5 + (\sqrt{1} + \cdots + \sqrt{8}) = \frac{2}{3}n^{3/2} + \frac{1}{2}\sqrt{n} + 0.194 \pm 0.0001$$

is guaranteed to never exceed $S(n)$ by more than 0.0834. For example, it gives the estimate ≈ 666716.473 for $S(10000)$.

Stirling's formula Let's use the trapezoidal rule to estimate the sum

$$F(n) = \ln 1 + \ln 2 + \cdots + \ln n.$$

Since the second derivative of $\ln x$ is $-1/x^2$, its value is between $-1/i^2$ and zero on every interval $[i, i+1]$, where $i \ge 1$. Inequality (6) gives that for every $i \ge 1$

$$\int_i^{i+1} \ln x \; dx = \frac{\ln i + \ln(i+1)}{2} + \varepsilon_i \quad \text{where} \quad 0 \le \varepsilon_i \le \frac{1}{3i^2}.$$

Therefore

$$\int_1^n \ln x \; dx = F(n) - \frac{1}{2}\ln n + E(n)$$

where

$$E(n) = \varepsilon_1 + \cdots + \varepsilon_n \le \frac{1}{3 \cdot 1^2} + \frac{1}{3 \cdot 2^2} + \cdots + \frac{1}{3 \cdot n^2}.$$

The antiderivative of $\ln x$ is $x \ln x - x + 1$, so

$$F(n) = n \ln n - n + \frac{1}{2}\ln n + 1 - E(n).$$

Exponentiating both sides and rearranging terms gives

$$e^{F(n)} = e^{n \ln n} \cdot e^{-n} \cdot e^{(\ln n)/2} \cdot e^{1-E(n)} = e^{1-E(n)} \cdot \sqrt{n} \cdot \left(\frac{n}{e}\right)^n.$$

On the other hand, $e^{F(n)} = e^{\ln 1 + \cdots + \ln n} = n!$. Using the integral method to upper bound $E(n)$ by $2/3$ we can conclude that

$$e^{1/3} \cdot \sqrt{n} \cdot \left(\frac{n}{e}\right)^n \le n! \le e \cdot \sqrt{n} \cdot \left(\frac{n}{e}\right)^n.$$

so that $n! = \Theta(\sqrt{n} \cdot (n/e)^n)$. In fact, we can say more. Since $E(n)$ is a sum of positive numbers $\varepsilon_1, \varepsilon_2, \ldots$ which is at most $2/3$, it must converge to some constant c as n approaches infinity. Therefore it must be that the expression

$$n! \Big/ \sqrt{n} \cdot \left(\frac{n}{e}\right)^n$$

approaches the constant $C = e^{1-c}$ as n approaches infinity. Using the law of large numbers for the binomial distribution from probability theory and a bit more calculus, one can show that in fact $C = \sqrt{2\pi}$ and deduce the following theorem:

Theorem 8 (Stirling's formula)**.** $\lim_{n\to\infty} n! \;/\; \sqrt{2\pi n} \cdot (n/e)^n = 1.$

References

This lecture is based on Chapter 12 of the text *Mathematics for Computer Science* by E. Lehman, T. Leighton, and A. Meyer. The variant of the integral methods described in the textbook is slightly different from the one in these notes.

Surprisingly, if we allow for the blocks to be stacked not only on top of one another but also side by side, the overhang can be much improved. If you are interested, see the amazing work *Overhang* by Mike Paterson and Uri Zwick.