Algorithms for Identifying Technical Patterns on Japanese Candle Stick Charts

Vittorio Papandrea

March 30, 2020

Abstract

This paper attempts to quantify, and form mathematical formulas along with algorithms to find technical patterns used by stock traders and economists. The data feed used is provided by IEX, all charts used contain minute candle data, and all price data is in $\mathbb{Z}$ using fixed point math instead of floating point math. These algorithms and math types are impleminted in the program riski. Riski gets its stock information from the IEX exchange.

1 Linear Equations in $\mathbb{Z}$

As said in the abstract, all math done is with integer arithmatic. This will generate a few challenges down the line because $\mathbb{Z}$ is not closed under division. A later more complicated analysis will require a linear equation whos slope is not zero which does not play well with integers. Below is a different way of representing a linear equation to be stable under the integers.

Proof. Consider the two points $(a, b), (c, d) \in \mathbb{Z}^2$ where $a, b, c, d \in \mathbb{Z}$ and $a \neq c$. Traditionally a linear equation is defined as $y = mx + b$ where $m = \frac{\Delta y}{\Delta x}$ and b is the intercept. This can pose problems for us because $\frac{d-b}{c-a} \in \mathbb{Q}$. For a generic linear equation made up of the two points $(a, b), (c, d)$ to be stable first we must show $f(a) = b$ and $f(c) = d$ which is not true for $y = mx + b$ because $m \notin \mathbb{Z}$ as shown above. Because of this we can not define our linear equation by slope and intercept and must find a way to stay in $\mathbb{Z}$ and satisfy $f(a) = b$ and $f(c) = d$ at the same time. With the help of algebra we can find a more generic form of a linear equation using the two generic points from above.

$$\begin{aligned}
y &= mx + b \\
y &= \frac{d-b}{c-a}x + b \\
y &= \frac{d-b}{c-a}x + (y_2 - (mx_2)) \quad \text{could be } y_1, x_1 \text{ if you wish} \\
y &= \frac{d-b}{c-a}x + (d - (\frac{d-b}{c-a}c)) \\
y &= \frac{d-b}{c-a}x + d - \frac{d-b}{c-a}c \\
y - d &= \frac{d-b}{c-a}x - \frac{d-b}{c-a}c \\
y - d &= (\frac{d-b}{c-a})(x - c) \\
y &= \frac{(d-b)(x-c)}{c-a} + d
\end{aligned} \tag{1}$$

With the new form of $y = f(x)$ we must show that the evaluation of $f(a) = b$ and $f(c) = d$ without the need of rounding. It is also worth to note for the curious that the y intercept b seems to have turned into d and while the slope m still holds x is shifted by c to the right. We will now use proof by cases. First we show the simpler case $f(c) = d$. Simply plugging it in we get

$$d = \frac{(d-b)(0)}{c-a} + d = d \tag{2}$$

That is stable since the fraction evaluates to $0 \in \mathbb{Z}$ and $d \in \mathbb{Z}$. Unfortunatly for the case $f(a) = b$ is it not so trivial. We will begin by plugging in a and b.

$$\begin{aligned} b &= \frac{(d-b)(a-c)}{c-a} + d \\ b - d &= \frac{(d-b)(a-c)}{c-a} \end{aligned} \tag{3}$$

Before continuing let us note that $b - d \in \mathbb{Z}$. Performing some cleaver factoring on the right hand side will allow us to eliminate some variables.

$$\begin{aligned} \frac{(d-b)(a-c)}{c-a} &\implies \frac{da - dc - ba + bc}{c-a} &\implies \\ \frac{d(a-c) + b(c-a)}{c-a} &\implies \frac{d(a-c)}{c-a} + \frac{b(c-a)}{c-a} &\implies \\ b - d &= \frac{d(a-c)}{c-a} + b \\ -d &= d\frac{(a-c)}{c-a} \end{aligned} \tag{4}$$

This is looking promising, all of the operations done above are just simplification and reording which means all operations have kept us in $\mathbb{Z}$. To stay in $\mathbb{Z}$ equation (4) must hold. Since $d, -d \in \mathbb{Z}$ we must show that for any $a, c \in \mathbb{Z} \quad a \neq c$, $\frac{a-c}{c-a} = -1$. This is trivial.

$$\begin{aligned} 1 &= 1 \\ \frac{c-a}{c-a} &= 1 \\ (-1)\frac{c-a}{c-a} &= 1(-1) \\ \frac{a-c}{c-a} &= -1 \end{aligned} \tag{5}$$

Therefore $d = d$. By the proof of both cases it is safe to say that

$$y = \frac{(d-b)(x-c)}{c-a} + d$$

is closed in $\mathbb{Z}$ for the two points $(a, b), (c, d)$.

But now we must do something about all points. This will require clever usage of ceil and floor. For our purposes, if the lines slope is positive we will take the floor and if the slope is negative we will take the ceil. This can we written as

$$y = \begin{cases} \left\lfloor \frac{(d-b)(x-c)}{c-a} \right\rfloor + d & \frac{d-b}{c-a} > 0 \\ \\ \left\lceil \frac{(d-b)(x-c)}{c-a} \right\rceil + d & \frac{d-b}{c-a} < 0 \\ \\ 0 & \text{undefined} \end{cases}$$

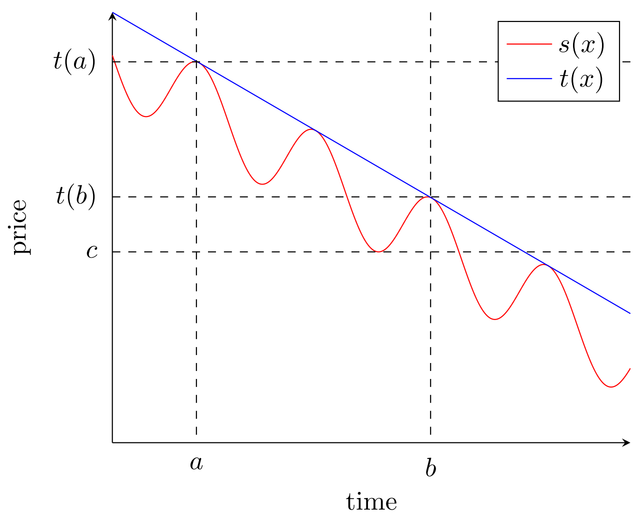
By flooring and ceiling the function we have restricted our linear equations range to $\mathbb{Z}$ and even without flooring or ceiling we are guanteed to produce integer results for the points $(a, b), (c, d)$. This is the most desirable way of evaluating linear equations while staying in $\mathbb{Z}$ for analysing japanese candle stick charts.

$\square$

2 Sloped Resistance and Support Lines

Sloped resistance and support lines are not at all easy to detect, Linear Equations in $\mathbb{Z}$ (1) is only used to represent the line, now one must construct and somehow quantify if it is a sloped resistance or support line. Throughout the remaining content of this section, downward sloped resistance lines will be used in mathematical and pseudocode representation, it will be easy to see how the same ideas and minimal code changes will result in generic sloped support lines.

Now that a trend line has been constructed we must find a way to quantify wheather or not the trend line is "good". We start with a theoretical model of a chart and trend line. Let us consider two functions, $t(x)$, representing a trend line and $s(x)$, representing the close values of a candle stick chart with $x \in \{[a, b] | a < b \vee a, b \in \mathbb{Z}\}$. Since $t(x)$ represents a resistance trend line, it is safe to assume $t(x) \geq s(x)$. Also, because $t(x)$ is downward sloping $max\ t(x) = t(a)$ and $min\ t(x) = t(b)$. Lastly $min\ s(x) = c$ where c is some constant. Below is a chart visualizing these theoretical constraints.



To measure how "good" or "bad" the trend is, it is not suffitient enough to just count the number of confirmations or spread the trend line covers, instead after building the trend line we will assign a score to the trend, the closer the number is to 0 the "better" the line. The closer it is to the upper bound, which will be derived below, will tell us its not the greatest line eventhough it meets the critiera to be a trend line.

* This idea came to mind while reading about the Lorenz curve and the Gini coefficient.

The score will be defined as how close $s(x)$ is to $t(x)$ this will be done by taking the area between these curves. First will make the wrong assumption that both $t(x)$ and $s(x)$ are continous in $\mathbb{Z}$ and then convert the integrals into their descrete equivilent. Now we find the upper and lower bounds of the area.

Proof. By definition $t(x) \geq s(x)$ therefore the area between $t(x)$ and $s(x)$

$$\int_a^b t(x) - s(x)\ dx \tag{6}$$

The minimum area between these curves is when $t(x) = s(x)$.

$$0 \leq \int_a^b t(x) - s(x)\ dx \tag{7}$$

The upper bound is a little more involved. The maximum area is when $s(x) = c$. In other words the minimum value of $s(x)$ is the entire curve. So,

$$0 \leq \int_a^b t(x) - s(x)\ dx \leq \int_a^b t(x) - c\ dx \tag{8}$$

It will be easier to expand these integrals first before finding their discrete equivilent.

$$0 \leq \int_a^b t(x)\ dx - \int_a^b s(x)\ dx \leq \int_a^b t(x)\ dx - \int_a^b c\ dx \tag{9}$$

Taking it one integral at a time $\int_a^b c\ dx$ is just the area of a rectangle therefore

$$\int_a^b c\ dx = c(b-a) \tag{10}$$

In the theoretical representation above it is clear that $\int_a^b t(x)\ dx$ is the area of a trapezoid. Therefore

$$\int_a^b t(x)\ dx = (b-a)(t(b) + \frac{1}{2}[t(a) - t(b)]) \tag{11}$$

Lastly by far the hardest integral, $\int_a^b s(x)\ dx$. If this were a continouse the function the trapezoidal rule would not give an exact integral, but since $s(x)$ is only defined in $\mathbb{Z}$ the trapezoidal rule applyed to $s(x)$ will give the exact integral. Therefore

$$\int_a^b s(x)\ dx = \frac{1}{2}[s(a) + \sum_{k=a+1}^{b-1}(2s(k)) + s(b)] \tag{12}$$

Putting it all together along with some simplification gives us bounds of

$$\begin{aligned} &0 \leq \\ &((b-a)(t(b) + \frac{1}{2}[t(a) - t(b)]) - (\frac{1}{2}[s(a) + \sum_{k=a+1}^{b-1}(2s(k)) + s(b)]) \leq \\ &(b-a)(t(b) + \frac{1}{2}[t(a) - t(b)] - c) \end{aligned} \tag{13}$$

□

Now that bounds have been established, the score of a trend line, defined as $f = \int_a^b t(x) - s(x)\ dx$, can be safly normalized between $[0, 1]$ and a user defined threshold g can be used to filter unwanted trend lines only allowing trend lines when $f \leq g$ to be considered valid.

3 Generic Candle Patterns

Before defining algorithms, we need to define mathematicaly the patterns traders look for, there are many types of patterns and we will only focus on the patterns that are currently implemented in riski.

3.1 Single Candle Patterns

Definition 1. Candle A candle is described as four points; $a, b, c, d \in \mathbb{Z}$ such that a is the top of the wick, b is the bottom of the wick, c is the top of the body, and d is the bottom of the body. To prevent flipping of any numbers, it is required that $a \geq c \geq d \geq b$.

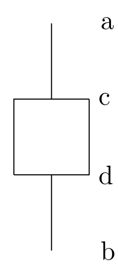


Figure 1: A generic candle

Definition 2. Marubozu A marubozu is identified as candle that has no wick. Consider a candle c defined by $a, b, c, d \in \mathbb{Z}$ then candle c is a marubozu iff $a = c$ and $d = b$ and $c \neq d$

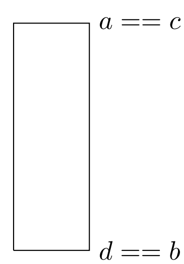


Figure 2: A generic marubozu

Definition 3. Spinning Top A spinning top is identified as candle whos ratios between body and wick are perfectly balanced. Consider a candle c defined by $a, b, c, d \in \mathbb{Z}$ then candle c is a spinning top iff $|a - c| = |c - d| = |d - b|$ and $a \neq b \neq c \neq d$

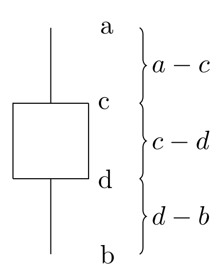


Figure 3: A generic spinning top