Minimal distance to a cubic function

Martin Thoma

December 11, 2013

When you want to develop a selfdriving car, you have to plan which path it should take. A reasonable choice for the representation of paths are cubic splines. You also have to be able to calculate how to steer to get or to remain on a path. A way to do this is applying the PID algorithm. This algorithm needs to know the signed current error. So you need to be able to get the minimal distance of a point to a cubic spline combined with the direction (left or right). As you need to get the signed error (and one steering direction might be prefered), it is not only necessary to get the minimal absolute distance, but might also help to get all points on the spline with minimal distance.

In this paper I want to discuss how to find all points on a cubic function with minimal distance to a given point. As other representations of paths might be easier to understand and to implement, I will also cover the problem of finding the minimal distance of a point to a polynomial of degree 0, 1 and 2.

1 Description of the Problem

Let $f : \mathbb{R} \to \mathbb{R}$ be a polynomial function and $P \in \mathbb{R}^2$ be a point. Let $d_{P,f} : \mathbb{R} \to \mathbb{R}_0^+$ be the Euklidean distance of a point P and a point $(x, f(x))$ on the graph of f:

$$d_{P,f}(x) := \sqrt{(x_P - x)^2 + (y_P - f(x))^2}$$

Now there is finite set $M = \{ x_1, \dots, x_n \}$ of minima for given f and P:

$$M = \left\{ x \in \mathbb{R} \;\middle|\; d_{P,f}(x) = \min_{\overline{x} \in \mathbb{R}} d_{P,f}(\overline{x}) \right\}$$

But minimizing $d_{P,f}$ is the same as minimizing $d_{P,f}^2$:

$$\begin{aligned} d_{P,f}(x)^2 &= \sqrt{(x_P - x)^2 + (y_P - f(x))^2}^2 && (1)\\ &= x_p^2 - 2x_p x + x^2 + y_p^2 - 2y_p f(x) + f(x)^2 && (2) \end{aligned}$$

Theorem 1 (Fermat's theorem about stationary points)
Let x_0 be a local extremum of a differentiable function $f : \mathbb{R} \to \mathbb{R}$.

Then: $f'(x_0) = 0$.

2 Minimal distance to a constant function

Let $f(x) = c$ with $c \in \mathbb{R}$ be a constant function.

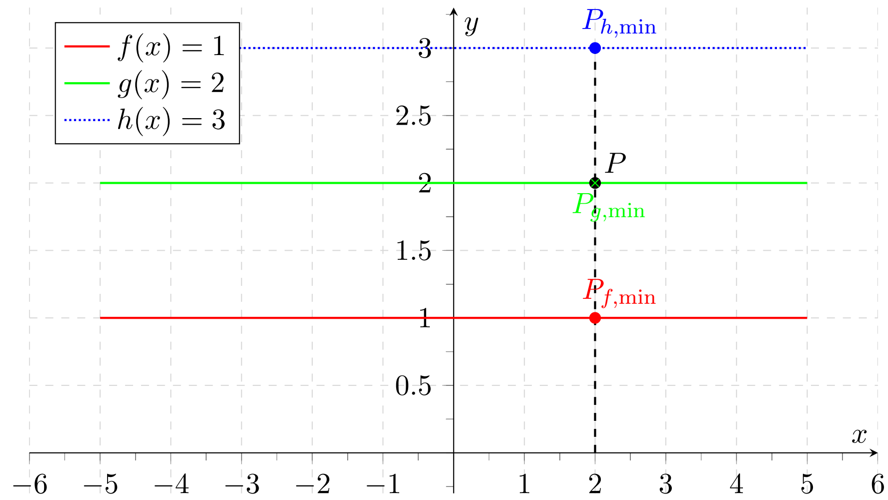

Figure 1: Three constant functions and their points with minimal distance

Then $(x_P, f(x_P))$ has minimal distance to P. Every other point has higher distance. See Figure 1.

3 Minimal distance to a linear function

Let $f(x) = m \cdot x + t$ with $m \in \mathbb{R} \setminus \{ 0 \}$ and $t \in \mathbb{R}$ be a linear function.

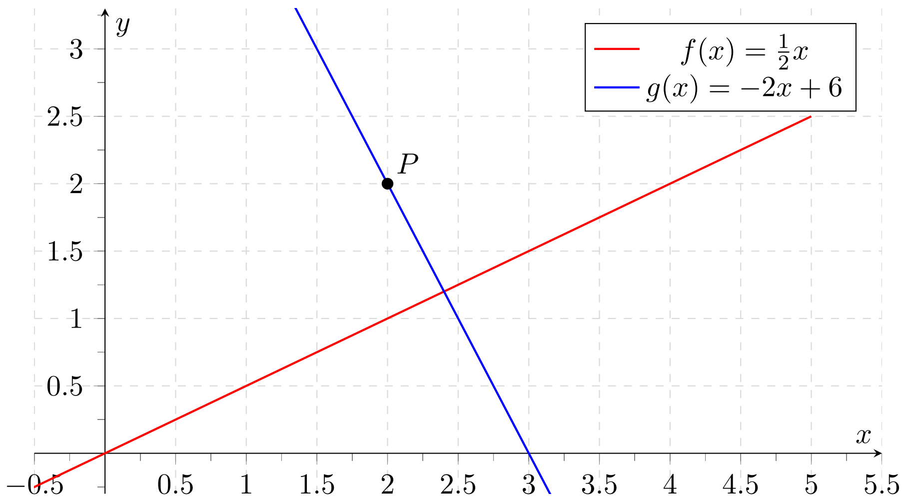

Figure 2: The shortest distance of P to f can be calculated by using the perpendicular

Now you can drop a perpendicular $f_\perp$ through P on $f(x)$. The slope of $f_\perp$ is $-\frac{1}{m}$ and $t_\perp$ can be calculated:

$$f_\perp(x) = -\frac{1}{m} \cdot x + t_\perp \tag{3}$$
$$\Rightarrow y_P = -\frac{1}{m} \cdot x_P + t_\perp \tag{4}$$
$$\Leftrightarrow t_\perp = y_P + \frac{1}{m} \cdot x_P \tag{5}$$

The point $(x, f(x))$ where the perpendicular $f_\perp$ crosses f is calculated this way:

$$f(x) = f_\perp(x) \tag{6}$$
$$\Leftrightarrow m \cdot x + t = -\frac{1}{m} \cdot x + \left(y_P + \frac{1}{m} \cdot x_P\right) \tag{7}$$
$$\Leftrightarrow \left(m + \frac{1}{m}\right) \cdot x = y_P + \frac{1}{m} \cdot x_P - t \tag{8}$$
$$\Leftrightarrow x = \frac{m}{m^2+1} \left(y_P + \frac{1}{m} \cdot x_P - t\right) \tag{9}$$

There is only one point with minimal distance. See Figure 2.

4 Minimal distance to a quadratic function

Let $f(x) = a \cdot x^2 + b \cdot x + c$ with $a \in \mathbb{R} \setminus \{ 0 \}$ and $b, c \in \mathbb{R}$ be a quadratic function.

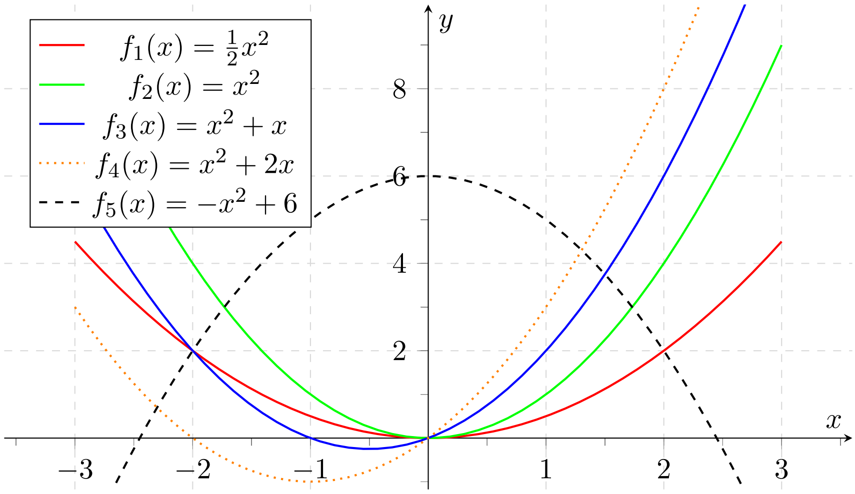

Figure 3: Quadratic functions

4.1 Calculate points with minimal distance

In this case, $d_{P,f}^2$ is polynomial of degree 4. We use Theorem 1:

$$
\begin{aligned}
0 &\stackrel{!}{=} (d_{P,f}^2)' && (10)\\
&= -2x_p + 2x - 2y_p f'(x) + \big(f(x)^2\big)' && (11)\\
&= -2x_p + 2x - 2y_p f'(x) + 2f(x) \cdot f'(x) \qquad \text{(chain rule)} && (12)\\
\Leftrightarrow 0 &\stackrel{!}{=} -x_p + x - y_p f'(x) + f(x) \cdot f'(x) \qquad \text{(divide by 2)} && (13)\\
&= -x_p + x - y_p(2ax + b) + (ax^2 + bx + c)(2ax + b) && (14)\\
&= -x_p + x - y_p \cdot 2ax - y_p b + (2a^2x^3 + 2abx^2 + 2acx + abx^2 + b^2x + bc) && (15)\\
&= -x_p + x - 2y_p ax - y_p b + (2a^2x^3 + 3abx^2 + 2acx + b^2x + bc) && (16)\\
&= 2a^2x^3 + 3abx^2 + (1 - 2y_p a + 2ac + b^2)x + (bc - by_p - x_p) && (17)
\end{aligned}
$$

This is an algebraic equation of degree 3. There can be up to 3 solutions in such an equation. Those solutions can be found with a closed formula.

Where are those closed formulas?

Example 1

Let $a = 1, b = 0, c = 1, x_p = 0, y_p = 1$. So $f(x) = x^2 + 1$ and $P(0, 1)$.

$$
\begin{aligned}
0 &\stackrel{!}{=} 4x^3 - 2x && (18)\\
&= 2x(2x^2 - 1) && (19)\\
\Rightarrow x_1 = 0 \quad x_{2,3} &= \pm\frac{1}{\sqrt{2}} && (20)
\end{aligned}
$$

As you can easily verify, only x_1 is a minimum of $d_{P,f}$.

4.2 Number of points with minimal distance

Theorem 2
A point P has either one or two points on the graph of a quadratic function f that are closest to P.

In the following, I will do some transformations with $f = f_0$ and $P = P_0$.

Moving f_0 and P_0 simultaneously in x or y direction does not change the minimum distance. Furthermore, we can find the points with minimum distance on the moved situation and calculate the minimum points in the original situation.

First of all, we move f_0 and P_0 by $\frac{b}{2a}$ in x direction, so

$$f_1(x) = ax^2 - \frac{b^2}{4a} + c \quad \text{and} \quad P_1 = \left(x_p + \frac{b}{2a},\ y_p\right)$$

Because:[1]

$$\begin{aligned} f(x - {}^{b}/_{2a}) &= a(x - {}^{b}/_{2a})^2 + b(x - {}^{b}/_{2a}) + c && (21)\\ &= a(x^2 - {}^{b}/_{a}x + {}^{b^2}/_{4a^2}) + bx - {}^{b^2}/_{2a} + c && (22)\\ &= ax^2 - bx + {}^{b^2}/_{4a} + bx - {}^{b^2}/_{2a} + c && (23)\\ &= ax^2 - {}^{b^2}/_{4a} + c && (24)\end{aligned}$$

Then move f_1 and P_1 by $\frac{b^2}{4a} - c$ in y direction. You get:

$$f_2(x) = ax^2 \quad \text{and} \quad P_2 = \Bigg(\underbrace{x_P + \frac{b}{2a}}_{=:z},\ \underbrace{y_P + \frac{b^2}{4a} - c}_{=:w}\Bigg)$$

Case 1: As $f_2(x) = ax^2$ is symmetric to the y axis, only points $P = (0, w)$ could possilby have three minima.

Then compute:

$$\begin{aligned} d_{P,f_2}(x) &= \sqrt{(x-0)^2 + (f_2(x) - w)^2} && (25)\\ &= \sqrt{x^2 + (ax^2 - w)^2} && (26)\\ &= \sqrt{x^2 + a^2x^4 - 2awx^2 + w^2} && (27)\\ &= \sqrt{a^2x^4 + (1-2aw)x^2 + w^2} && (28)\\ &= \sqrt{\left(a^2x^2 + \frac{1-2aw}{2}\right)^2 + w^2 - (1-2aw)^2} && (29)\\ &= \sqrt{\left(a^2x^2 + {}^{1}/_{2} - aw\right)^2 + \left(w^2 - (1-2aw)^2\right)} && (30)\end{aligned}$$

The term

$$a^2x^2 + ({}^{1}/_{2} - aw)$$

[1] The idea why you subtract $\frac{b}{2a}$ within f is that when you subtract something from x before applying f it takes more time (x needs to be bigger) to get to the same situation. So to move the whole graph by 1 to the left whe have to add +1.

should get as close to 0 as possilbe when we want to minimize d_{P,f_2}. For $w \leq 1/2a$ you only have $x = 0$ as a minimum. For all other points $P = (0, w)$, there are exactly two minima $x_{1,2} = \pm\sqrt{aw - 1/2}$.

Case 2: $P = (z, w)$ is not on the symmetry axis, so $z \neq 0$. Then you compute:

$$
\begin{aligned}
d_{P,f_2}(x) &= \sqrt{(x-z)^2 + (f(x) - w)^2} && (31)\\
&= \sqrt{(x^2 - 2zx + z^2) + ((ax^2)^2 - 2awx^2 + w^2)} && (32)\\
&= \sqrt{a^2x^4 + (1-2aw)x^2 + (-2z)x + z^2 + w^2} && (33)\\
0 &\stackrel{!}{=} \left(\left(d_{P,f_2}(x)\right)^2\right)' && (34)\\
&= 4a^2x^3 + 2(1-2aw)x + (-2z) && (35)\\
&= 2\left(2a^2x^2 + (1-2aw)\right)x - 2z && (36)\\
\Leftrightarrow 0 &\stackrel{!}{=} (2a^2x^2 + (1-2aw))x - z && (37)\\
&= 2a^2x^3 + (1-2aw)x - z && (38)\\
\Leftrightarrow 0 &\stackrel{!}{=} x^3 + \underbrace{\frac{(1-2aw)}{2a^2}}_{=:\alpha}x + \underbrace{\frac{-z}{2a^2}}_{=:\beta} && (39)\\
&= x^3 + \alpha x + \beta && (40)
\end{aligned}
$$

The solution of Equation 40 is

$$t := \sqrt[3]{\sqrt{3 \cdot (4a^3 + 27b^2)} - 9b}$$

$$x = \frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}a}{t}$$

verify this solution

So the solution is given by

$$x_S := -\frac{b}{2a} \quad \text{(the symmetry axis)}$$

$$\arg\min_{x \in \mathbb{R}} d_{P,f}(x) = \begin{cases} x_1 = +\sqrt{a(y_p + \frac{b^2}{4a} - c) - \frac{1}{2}} + x_S \text{ and} & \text{if } x_P = x_S \text{ and } y_p + \frac{b^2}{4a} - c > \frac{1}{2a} \\ x_2 = -\sqrt{a(y_p + \frac{b^2}{4a} - c) - \frac{1}{2}} + x_S & \\ x_1 = x_S & \text{if } x_P = x_S \text{ and } y_p + \frac{b^2}{4a} - c \leq \frac{1}{2a} \\ x_1 = todo & \text{if } x_P \neq x_S \end{cases}$$

5 Minimal distance to a cubic function

Let $f(x) = a \cdot x^3 + b \cdot x^2 + c \cdot x + d$ be a cubic function with $a \in \mathbb{R} \setminus \{0\}$ and $b, c, d \in \mathbb{R}$ be a function.

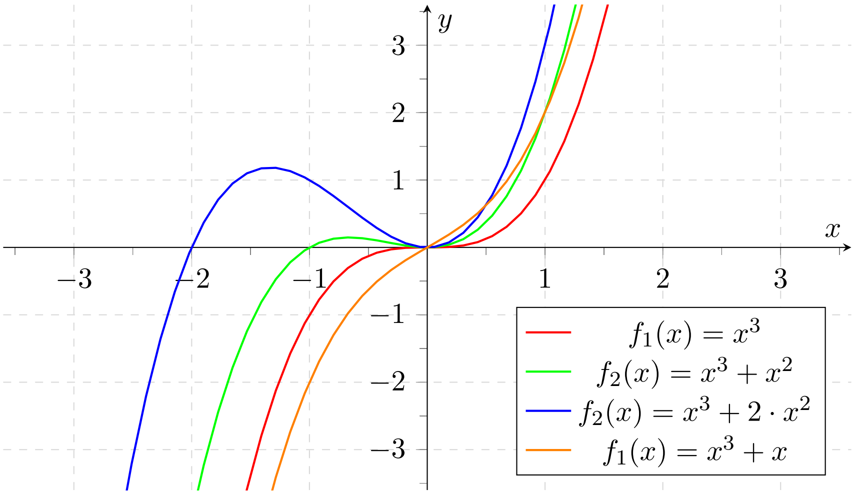

Figure 4: Cubic functions

5.1 Calculate points with minimal distance

Theorem 3
There cannot be an algebraic solution to the problem of finding a closest point $(x, f(x))$ to a given point P when f is a polynomial function of degree 3 or higher.

Proof: Let $g : \mathbb{R} \to \mathbb{R}$ be a polynomial of degree 5

$$g(x) = \tilde{a}x^5 + \tilde{b}x^4 + \tilde{c}x^3 + \tilde{d}x^2 + \tilde{e}x + \tilde{f}$$

with $\tilde{a} \in \mathbb{R}_{>0}$, $\tilde{b} \in \mathbb{R} \setminus \{0\}$ and $\tilde{c}, \tilde{d}, \tilde{e}, \tilde{f} \in \mathbb{R}$. Then, according to the Abel-Ruffini theorem, the equation

$$g(x) = 0$$

cannot be solved algebraicly.

So you can find a, b, c, d, x_p, y_p such that

$$g(x) = \underbrace{3a^2}_{=\tilde{a}} x^5 + \underbrace{5ab}_{=\tilde{b}} x^4 + \underbrace{2(2ac + b^2)}_{=\tilde{c}} x^3 \quad + \underbrace{3(ad + bc - ay_p)}_{=\tilde{d}} x^2 \tag{41}$$

$$+ \underbrace{(2bd + c^2 + 1 - 2by_p)}_{=\tilde{e}} x + \underbrace{cd - cy_p - x_p}_{=\tilde{f}} \tag{42}$$

$$= f'(x) \cdot (f(x) - y_p) + (x - x_p) \tag{43}$$

$$= f(x) \cdot f'(x) - y_p f'(x) + x - x_p \tag{44}$$

And

$$
\begin{aligned}
& g(x) \stackrel{!}{=} 0 && (45)\\
\Leftrightarrow\ & 0 \stackrel{!}{=} 2f(x)\cdot f'(x) - 2y_p f'(x) + 2x - 2x_p && (46)\\
& = -2x_p + 2x - 2y_p(f(x))' + (f(x)^2)' && (47)\\
& = \left((x-x_p)^2\right)' + \left((f(x)-y_p)^2\right)' && (48)\\
& = \left((x-x_p)^2 + (f(x)-y_p)^2\right)' && (49)\\
& = (d_{P,f}(x)^2)' && (50)
\end{aligned}
$$

So the problem of finding a closest point $(x, f(x))$ on a cubic function f to P is essentially the same as finding a root of a polynomial function of degree 5. As this cannot be solved algebraicly, the problem of finding such a point can also not be solved algebraicly. ■

Start with theorem that this problem is not solvable with analytics only. Use a general 5th degree function and show that it can be mapped to a f and P instance.

When you want to calculate points with minimal distance, you can take the same approach as in Equation 13:

$$
\begin{aligned}
0 & \stackrel{!}{=} -2x_p + 2x - 2y_p(f(x))' + (f(x)^2)' && (51)\\
& = 2f(x)\cdot f'(x) - 2y_p f'(x) + 2x - 2x_p && (52)\\
& = f(x)\cdot f'(x) - y_p f'(x) + x - x_p && (53)\\
& = \underbrace{f'(x)\cdot(f(x)-y_p)}_{\text{Polynomial of degree 5}} + x - x_p && (54)
\end{aligned}
$$

General algebraic equations of degree 5 don't have a solution formula.[2] Although here seems to be more structure, the resulting algebraic equation can be almost any polynomial of degree 5:[3]

$$
\begin{aligned}
0 & \stackrel{!}{=} f'(x)\cdot(f(x)-y_p) + (x-x_p) && (55)\\
& = \underbrace{3a^2}_{=\tilde{a}} x^5 + \underbrace{5ab}_{\tilde{b}} x^4 + \underbrace{2(2ac+b^2)}_{=:\tilde{c}} x^3 \quad + \underbrace{3(ad+bc-ay_p)}_{\tilde{d}} x^2 && (56)\\
& \quad + \underbrace{(2bd+c^2+1-2by_p)}_{=:\tilde{e}} x + \underbrace{cd - cy_p - x_p}_{=:\tilde{f}} && (57)\\
0 & \stackrel{!}{=} \tilde{a}x^5 + \tilde{b}x^4 + \tilde{c}x^3 + \tilde{d}x^2 + \tilde{e}x + \tilde{f} && (58)
\end{aligned}
$$

1. With a, we can get any value of $\tilde{a} \in \mathbb{R} \setminus \{0\}$.
2. With b, we can get any value of $\tilde{b} \in \mathbb{R} \setminus \{0\}$.
3. With c, we can get any value of $\tilde{c} \in \mathbb{R}$.
4. With d, we can get any value of $\tilde{d} \in \mathbb{R}$.
5. With y_p, we can get any value of $\tilde{e} \in \mathbb{R}$.

[2] TODO: Quelle
[3] Thanks to Peter Košinár on math.stackexchange.com for this one

6. With x_p, we can get any value of $\tilde{f} \in \mathbb{R}$.

The first restriction only guaratees that we have a polynomial of degree 5. The second one is necessary, to get a high range of $\tilde{e}$.

This means, that there is no solution formula for the problem of finding the closest points on a cubic function to a given point.

5.2 Another approach

Just like we moved the function f and the point to get in a nicer situation, we can apply this approach for cubic functions.

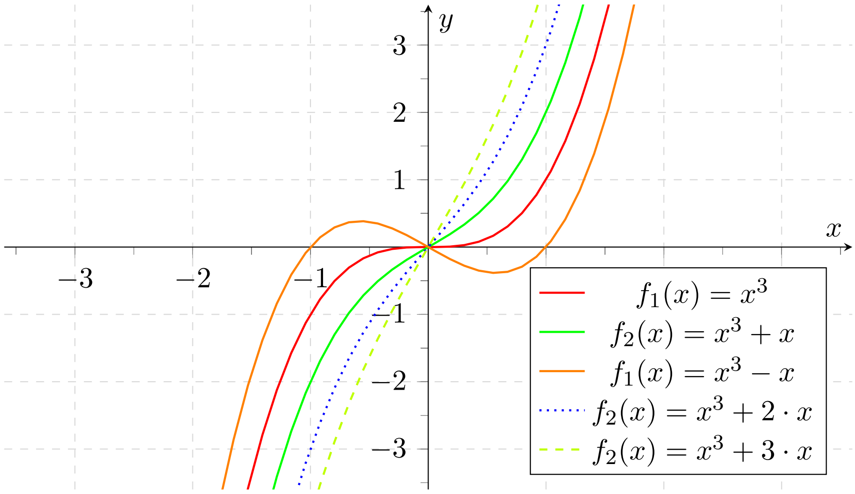

Figure 5: Cubic functions with $b = d = 0$

First, we move f_0 by $\frac{b}{3a}$ to the right, so

$$f_1(x) = ax^3 + \frac{b^2(c-1)}{3a}x + \frac{2b^3}{27a^2} - \frac{bc}{3a} + d \quad \text{and} \quad P_1 = (x_P + \frac{b}{3a}, y_P)$$

because

$$\begin{aligned}
f_1(x) &= a\left(x - \frac{b}{3a}\right)^3 + b\left(x - \frac{b}{3a}\right)^2 + c\left(x - \frac{b}{3a}\right) + d && (59)\\
&= a\left(x^3 - 3\frac{b}{3a}x^2 + 3(\frac{b}{3a})^2 x - \frac{b^3}{27a^3}\right) + b\left(x^2 - \frac{2b}{3a}x + \frac{b^2}{9a^2}\right) + cx - \frac{bc}{3a} + d && (60)\\
&= ax^3 - bx^2 + \frac{b^2}{3a}x - \frac{b^3}{27a^2} && (61)\\
&\quad + bx^2 - \frac{2b^2}{3a}x + \frac{b^3}{9a^2} && (62)\\
&\quad\quad + cx - \frac{bc}{3a} + d && (63)\\
&= ax^3 + \frac{b^2}{3a}(1 - 2 + c)\,x + \frac{b^3}{9a^2}\left(1 - \frac{1}{3}\right) - \frac{bc}{3a} + d && (64)
\end{aligned}$$

Which way to move might be clever?

5.3 Number of points with minimal distance

As there is an algebraic equation of degree 5, there cannot be more than 5 solutions.

Can there be 3, 4 or even 5 solutions? Examples!
After looking at function graphs of cubic functions, I'm pretty sure that there cannot be 4 or 5 solutions, no matter how you chose the cubic function f and P.
I'm also pretty sure that there is no polynomial (no matter what degree) that has more than 3 solutions.

6 Newtons method

When does Newtons method converge? How fast? How to choose starting point?

7 Quadratic minimization

TODO

8 Conclusion

TODO