Minimal distance to a cubic function

Martin Thoma

December 12, 2013

Introduction

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.

While I analyzed this problem, I've got interested in variations of the underlying PID-related problem. So I will try to give robust and easy-to-implement algorithms to calculated the distance of a point to a (piecewise or global) defined polynomial function of degree ≤ 3.

Contents

1 Description of the Problem

Let $f : D \to \mathbb{R}$ with $D \subseteq \mathbb{R}$ 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 $M = \{ x_1, \dots, x_n \} \subseteq D$ of minima for given f and P:

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

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

$$d_{P,f}(x)^2 = \sqrt{(x_P - x)^2 + (y_P - f(x))^2}^2 \tag{1.1}$$
$$= x_p^2 - 2x_p x + x^2 + y_p^2 - 2y_p f(x) + f(x)^2 \tag{1.2}$$

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 Constant functions

2.1 Defined on $\mathbb{R}$

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

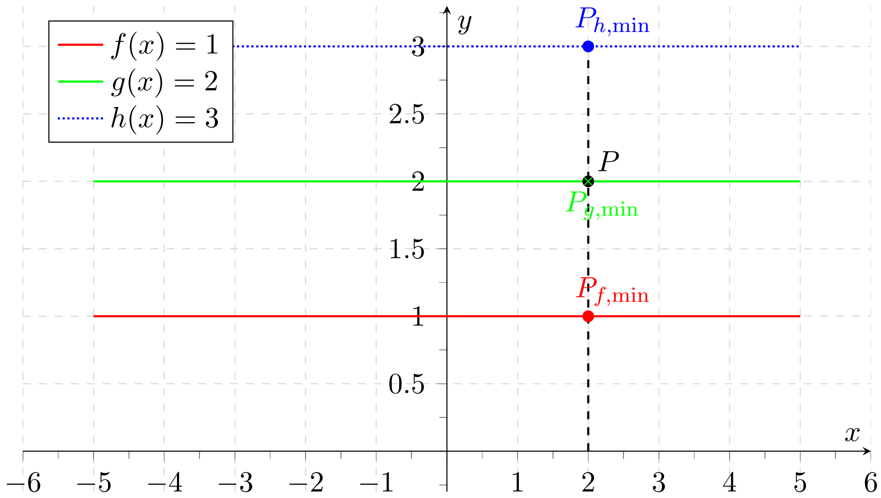

Figure 2.1: Three constant functions and their points with minimal distance

Then $(x_P, f(x_P))$ has minimal distance to P. Every other point has higher distance. See Figure 2.1.

2.2 Defined on a closed interval of $\mathbb{R}$

3 Linear function

3.1 Defined on $\mathbb{R}$

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

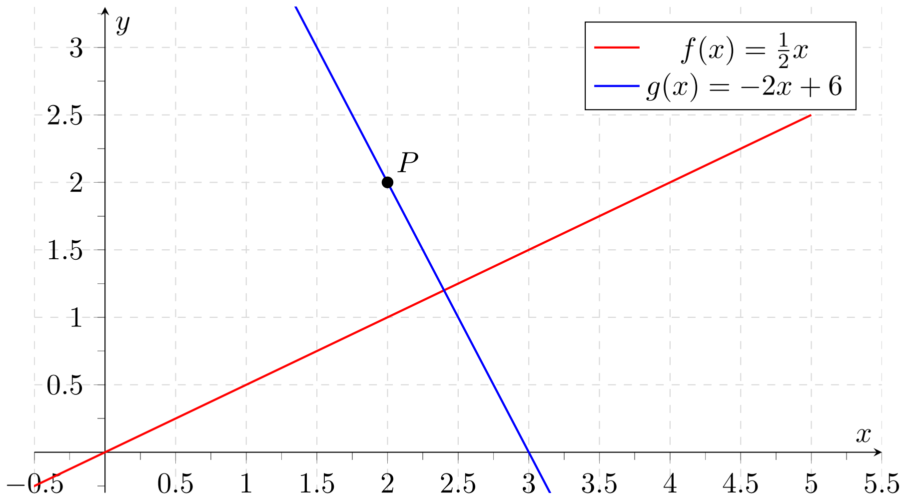

Figure 3.1: The shortest distance of P to f can be calculated by using the perpendicular

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:

$$f_\perp(x) = -\frac{1}{m} \cdot x + t_\perp \tag{3.1}$$

$$\Rightarrow y_P = -\frac{1}{m} \cdot x_P + t_\perp \tag{3.2}$$

$$\Leftrightarrow t_\perp = y_P + \frac{1}{m} \cdot x_P \tag{3.3}$$

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

$$f(x) = f_\perp(x) \tag{3.4}$$

$$\Leftrightarrow m \cdot x + t = -\frac{1}{m} \cdot x + \left(y_P + \frac{1}{m} \cdot x_P \right) \tag{3.5}$$

$$\Leftrightarrow \left(m + \frac{1}{m} \right) \cdot x = y_P + \frac{1}{m} \cdot x_P - t \tag{3.6}$$

$$\Leftrightarrow x = \frac{m}{m^2 + 1} \left(y_P + \frac{1}{m} \cdot x_P - t \right) \tag{3.7}$$

There is only one point with minimal distance. See Figure 3.1.

3.2 Defined on a closed interval of $\mathbb{R}$

4 Quadratic functions

4.1 Defined on $\mathbb{R}$

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.

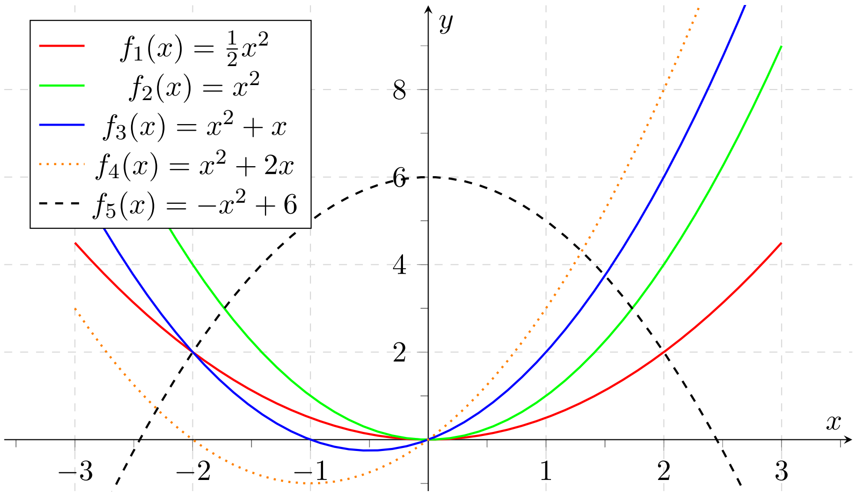

Figure 4.1: Quadratic functions

4.1.1 Calculate points with minimal distance

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

$$\begin{aligned}
0 &\stackrel{!}{=} (d_{P,f}^2)' && (4.1)\\
&= -2x_p + 2x - 2y_p f'(x) + \left(f(x)^2\right)' && (4.2)\\
&= -2x_p + 2x - 2y_p f'(x) + 2f(x) \cdot f'(x) \qquad \text{(chain rule)} && (4.3)\\
\Leftrightarrow 0 &\stackrel{!}{=} -x_p + x - y_p f'(x) + f(x) \cdot f'(x) \qquad \text{(divide by 2)} && (4.4)\\
&= -x_p + x - y_p(2ax + b) + (ax^2 + bx + c)(2ax + b) && (4.5)\\
&= -x_p + x - y_p \cdot 2ax - y_p b + (2a^2x^3 + 2abx^2 + 2acx + abx^2 + b^2x + bc) && (4.6)\\
&= -x_p + x - 2y_p ax - y_p b + (2a^2x^3 + 3abx^2 + 2acx + b^2x + bc) && (4.7)\\
&= 2a^2x^3 + 3abx^2 + (1 - 2y_p a + 2ac + b^2)x + (bc - by_p - x_p) && (4.8)
\end{aligned}$$

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 $P(0, 1)$.

$$0 \stackrel{!}{=} 4x^3 - 2x \tag{4.9}$$
$$= 2x(2x^2 - 1) \tag{4.10}$$
$$\Rightarrow x_1 = 0 \quad x_{2,3} = \pm\frac{1}{\sqrt{2}} \tag{4.11}$$

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

4.1.2 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]

$$f(x - {}^{b}/_{2a}) = a(x - {}^{b}/_{2a})^2 + b(x - {}^{b}/_{2a}) + c \tag{4.12}$$
$$= a(x^2 - {}^{b}/_{a}x + {}^{b^2}/_{4a^2}) + bx - {}^{b^2}/_{2a} + c \tag{4.13}$$
$$= ax^2 - bx + {}^{b^2}/_{4a} + bx - {}^{b^2}/_{2a} + c \tag{4.14}$$
$$= ax^2 - {}^{b^2}/_{4a} + c \tag{4.15}$$

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

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

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

[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 $+1$.

Then compute:

$$
\begin{aligned}
d_{P,f_2}(x) &= \sqrt{(x-0)^2 + (f_2(x) - w)^2} && (4.16)\\
&= \sqrt{x^2 + (ax^2 - w)^2} && (4.17)\\
&= \sqrt{x^2 + a^2x^4 - 2awx^2 + w^2} && (4.18)\\
&= \sqrt{a^2x^4 + (1-2aw)x^2 + w^2} && (4.19)\\
&= \sqrt{\left(a^2x^2 + \frac{1-2aw}{2}\right)^2 + w^2 - (1-2aw)^2} && (4.20)\\
&= \sqrt{\left(a^2x^2 + 1/2 - aw\right)^2 + \left(w^2 - (1-2aw)^2\right)} && (4.21)
\end{aligned}
$$

The term

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

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:

$$
\begin{aligned}
d_{P,f_2}(x) &= \sqrt{(x-z)^2 + (f(x) - w)^2} && (4.22)\\
&= \sqrt{(x^2 - 2zx + z^2) + ((ax^2)^2 - 2awx^2 + w^2)} && (4.23)\\
&= \sqrt{a^2x^4 + (1-2aw)x^2 + (-2z)x + z^2 + w^2} && (4.24)\\
0 &\stackrel{!}{=} \left(\left(d_{P,f_2}(x)\right)^2\right)' && (4.25)\\
&= 4a^2x^3 + 2(1-2aw)x + (-2z) && (4.26)\\
&= 2\left(2a^2x^2 + (1-2aw)\right)x - 2z && (4.27)\\
\Leftrightarrow 0 &\stackrel{!}{=} (2a^2x^2 + (1-2aw))x - z && (4.28)\\
&= 2a^2x^3 + (1-2aw)x - z && (4.29)\\
\Leftrightarrow 0 &\stackrel{!}{=} x^3 + \underbrace{\frac{(1-2aw)}{2a^2}}_{=:\alpha} x + \underbrace{\frac{-z}{2a^2}}_{=:\beta} && (4.30)\\
&= x^3 + \alpha x + \beta && (4.31)
\end{aligned}
$$

The solution of Equation 4.31 is

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

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

When you insert this in Equation 4.31 you get:[2]

$$0 \stackrel{!}{=} \left(\frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right)^3 + \alpha\left(\frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right) + \beta \qquad (4.32)$$

[2] Remember: $(a-b)^3 = a^3 - 3a^2b + 3ab^2 - b^3$

$$= \left(\frac{t}{\sqrt[3]{18}}\right)^3 - 3\left(\frac{t}{\sqrt[3]{18}}\right)^2 \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t} + 3\left(\frac{t}{\sqrt[3]{18}}\right)\left(\frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right)^2 - \left(\frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right)^3 + \alpha \left(\frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right) + \beta \quad (4.33)$$

$$= \frac{t^3}{18} - \frac{3t^2}{\sqrt[3]{18^2}} \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t} + \frac{3t}{\sqrt[3]{18}} \frac{\sqrt[3]{\frac{4}{9}}\alpha^2}{t^2} - \frac{\frac{2}{3}\alpha^3}{t^3} + \alpha \left(\frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right) + \beta \quad (4.34)$$

$$= \frac{t^3}{18} - \frac{\sqrt[3]{18}t\alpha}{\sqrt[3]{18^2}} + \frac{\sqrt[3]{12}\alpha^2}{\sqrt[3]{18}t} - \frac{\frac{2}{3}\alpha^3}{t^3} + \alpha \left(\frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right) + \beta \quad (4.35)$$

$$= \frac{t^3}{18} - \frac{t\alpha}{\sqrt[3]{18}} + \frac{\sqrt[3]{2}\alpha^2}{\sqrt[3]{3}t} - \frac{\frac{2}{3}\alpha^3}{t^3} + \alpha \left(\frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t}\right) + \beta \quad (4.36)$$

$$= \frac{t^3}{18} - \frac{t\alpha}{\sqrt[3]{18}} - \frac{\frac{2}{3}\alpha^3}{t^3} + \frac{\alpha t}{\sqrt[3]{18}} + \beta \quad (4.37)$$

$$= \frac{t^3}{18} - \frac{\frac{2}{3}\alpha^3}{t^3} + \beta \quad (4.38)$$

$$= \frac{t^6 - 12\alpha^3 + \beta 18t^3}{18t^3} \quad (4.39)$$

Now only go on calculating with the numerator. Start with resubstituting t:

$$0 = \left(\sqrt{3 \cdot (4\alpha^3 + 27\beta^2)} - 9\beta\right)^2 - 12\alpha^3 + \beta 18\left(\sqrt{3 \cdot (4\alpha^3 + 27\beta^2)} - 9\beta\right) \quad (4.40)$$

$$= \left(\sqrt{3 \cdot (4\alpha^3 + 27\beta^2)}\right)^2 + (9\beta)^2 - 12\alpha^3 - 18 \cdot 9\beta^2 \quad (4.41)$$

$$= 3 \cdot (4\alpha^3 + 27\beta^2) - 81\beta^2 - 12\alpha^3 \quad (4.42)$$

$$= (4\alpha^3 + 27\beta^2) - 27\beta^2 - 4\alpha^3 \quad (4.43)$$

$$= 0 \quad (4.44)$$

So the solution is given by

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

$$w := y_P + \frac{b^2}{4a} - c \quad \text{and} \quad z := x_P + \frac{b}{2a}$$

$$\alpha := \frac{(1 - 2aw)}{2a^2} \quad \text{and} \quad \beta := \frac{-z}{2a^2}$$

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

$$\arg\min_{x \in \mathbb{R}} d_{P,f}(x) = \begin{cases} x_1 = +\sqrt{a(y_p + \frac{b^2}{4a} - c) - \frac{1}{2}} + x_S \text{ and} & \text{if } x_P = x_S \text{ and } y_p + \frac{b^2}{4a} - c > \frac{1}{2a} \\ x_2 = -\sqrt{a(y_p + \frac{b^2}{4a} - c) - \frac{1}{2}} + x_S & \\ x_1 = x_S & \text{if } x_P = x_S \text{ and } y_p + \frac{b^2}{4a} - c \leq \frac{1}{2a} \\ x_1 = \frac{t}{\sqrt[3]{18}} - \frac{\sqrt[3]{\frac{2}{3}}\alpha}{t} & \text{if } x_P \neq x_S \end{cases}$$

4.2 Defined on a closed interval of $\mathbb{R}$

5 Cubic functions

5.1 Defined on $\mathbb{R}$

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.

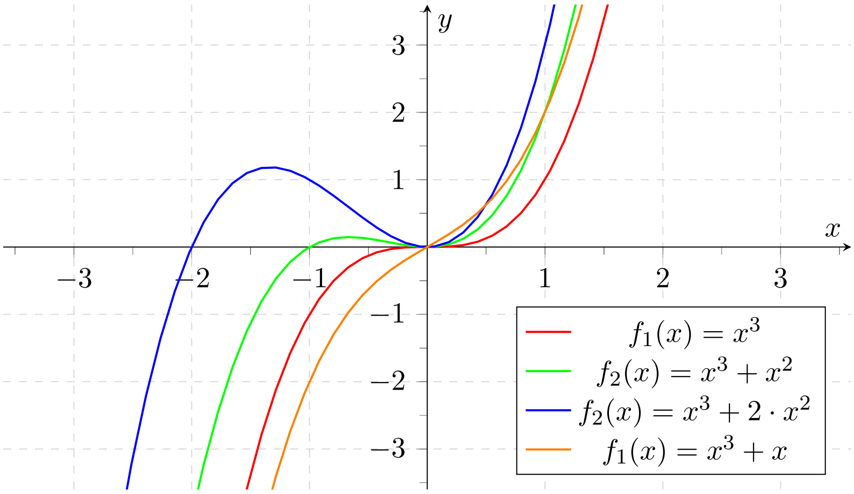

Figure 5.1: Cubic functions

5.1.1 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: Suppose you could solve the closest point problem for arbitrary cubic functions $f = ax^3 + bx^2 + cx + d$ and arbitrary points $P = (x_P, y_P)$.

Then you could solve the following problem for x:

$$0 \stackrel{!}{=} \left((d_{P,f}(x))^2\right)' \qquad = -2x_p + 2x - 2y_p(f(x))' + (f(x)^2)' \tag{5.1}$$

$$= 2f(x) \cdot f'(x) - 2y_p f'(x) + 2x - 2x_p \tag{5.2}$$

$$= f(x) \cdot f'(x) - y_p f'(x) + x - x_p \tag{5.3}$$

$$= \underbrace{f'(x) \cdot (f(x) - y_p)}_{\text{Polynomial of degree 5}} + x - x_p \tag{5.4}$$

General algebraic equations of degree 5 don't have a solution formula.[1] Although here seems

[1]TODO: Quelle

to be more structure, the resulting algebraic equation can be almost any polynomial of degree 5:[2]

$$
\begin{aligned}
0 &\stackrel{!}{=} f'(x) \cdot (f(x) - y_p) + (x - x_p) && (5.5)\\
&= \underbrace{3a^2}_{=\tilde{a}} x^5 + \underbrace{5ab}_{\tilde{b}} x^4 + \underbrace{2(2ac + b^2)}_{=:\tilde{c}} x^3 \quad + \underbrace{3(ad + bc - ay_p)}_{\tilde{d}} x^2 && (5.6)\\
&\qquad + \underbrace{(2bd + c^2 + 1 - 2by_p)}_{=:\tilde{e}} x + \underbrace{cd - cy_p - x_p}_{=:\tilde{f}} && (5.7)
\end{aligned}
$$

$$0 \stackrel{!}{=} \tilde{a}x^5 + \tilde{b}x^4 + \tilde{c}x^3 + \tilde{d}x^2 + \tilde{e}x + \tilde{f} \tag{5.8}$$

1. For any coefficient $\tilde{a} \in \mathbb{R}_{>0}$ of x^5 we can choose a such that we get $\tilde{a}$.
2. For any coefficient $\tilde{b} \in \mathbb{R} \setminus \{ 0 \}$ of x^4 we can choose b such that we get $\tilde{b}$.
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 $\tilde{e} \in \mathbb{R}$.
6. With x_p, we can get any value of $\tilde{f} \in \mathbb{R}$.

The first restriction guarantees 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, because if there was one, you could use this formula for finding roots of polynomials of degree 5. ■

5.1.2 Another approach

Currently, this is only an idea. It might be usefull to move the cubic function f such that f is point symmetric to the origin. But I'm not sure how to make use of this symmetry.

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

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

$$
\begin{aligned}
f_1(x) &= a\left(x - \frac{b}{3a}\right)^3 + b\left(x - \frac{b}{3a}\right)^2 + c\left(x - \frac{b}{3a}\right) + d && (5.9)\\
&= a\left(x^3 - 3\frac{b}{3a}x^2 + 3(\frac{b}{3a})^2 x - \frac{b^3}{27a^3}\right) + b\left(x^2 - \frac{2b}{3a}x + \frac{b^2}{9a^2}\right) + cx - \frac{bc}{3a} + d && (5.10)
\end{aligned}
$$

[2] Thanks to Peter Košinár on math.stackexchange.com for this one

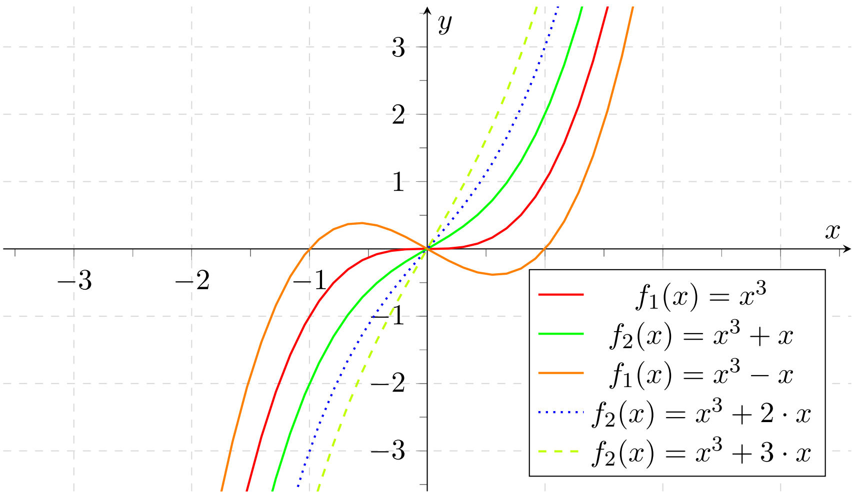

Figure 5.2: Cubic functions with $b = d = 0$

$$= ax^3 - bx^2 + \frac{b^2}{3a}x - \frac{b^3}{27a^2} \tag{5.11}$$

$$+ bx^2 - \frac{2b^2}{3a}x + \frac{b^3}{9a^2} \tag{5.12}$$

$$+ cx - \frac{bc}{3a} + d \tag{5.13}$$

$$= ax^3 + \frac{b^2}{3a}(1 - 2 + c)\, x + \frac{b^3}{9a^2}\left(1 - \frac{1}{3}\right) - \frac{bc}{3a} + d \tag{5.14}$$

5.1.3 Number of points with minimal distance

As this leads to a polynomial of degree 5 of which we have to find roots, 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.

5.1.4 Interpolation and approximation

Quadratic spline interpolation

You could interpolate the cubic function by a quadratic spline.

Bisection method

TODO

Newtons method

One way to find roots of functions is Newtons method. It gives an iterative computation procedure that can converge quadratically if some conditions are met:

Theorem 4 (local quadratic convergence of Newton's method)
Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^n \in C^2(\mathbb{R})$. Let $x^* \in D$ with $f(x^*) = 0$ and the Jaccobi-Matrix $f'(x^*)$ should not be invertable when evaluated at the root.

Then there is a sphere

$$K := K_\rho(x^*) = \{ x \in \mathbb{R}^n \mid \|x - x^*\|_\infty \leq \rho \} \subseteq D$$

such that x^* is the only root of f in K. Furthermore, the elements of the sequence

$$x_{n+1} = x_n - \frac{f'(x_n)}{f(x_n)}$$

are for every starting value $x_0 \in K$ again in K and

$$\lim_{n\to\infty} x_k = x^*$$

Also, there is a constant $C > 0$ such that

$$\|x^* - x_{n+1}\| = C\|x^* - x_n\|^2 \text{ for } n \in \mathbb{N}_0\|$$

The approach is extraordinary simple. You choose a starting value x_0 and compute

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

As soon as the values don't change much, you are close to a root. The problem of this approach is choosing a starting value that is close enough to the root. So we have to have a "good" initial guess.

Quadratic minimization

TODO

5.2 Defined on a closed interval of $\mathbb{R}$