www.journal.asbsociety.com

JASB
JOURNAL OF THE ASB SOCIETY

Optimizing length of planar curves

Kloud, Vojtěch,[1]* **Bednařík, Dušan**[2]

1 *První soukromé jazykove gymnazium, Brandlova 875, Hradec králové 500 03, Czechia,*
2 *University of Hradec Králové, Rokitanského 62, Hradec Králové 500 03, Czechia,*
* *Corresponding author: vojtech.kloud@gmail.com*

This article focuses on the problem of finding a shortest path in plane with obstacles. Problems of such nature occur for instance in robotics or transport and are of great importance. The problem is analyzed using the methods of mathematical analysis and calculus of variations. Definitions of basic concepts of the problem are given. From these definitions, useful properties, such as convexity of the length functional, are proven. These properties are used to show the existence of a solution in one of the considered cases of the problem. Other case of the problem was considered, where it is established under which conditions does a shortest path attain its general form and what this form looks like.

Key words: Optimization, calculus of variations, planar curves, constraints, obstacles.

(Submitted: 5 October 2020, Accepted: 12 October 2020, Published: 28 December 2020)

1 Introduction

The main goal of this paper is to answer a seemingly trivial question; what is the shortest path in a plane between two endpoints that avoids given obstacles? People usually have a strong intuition for finding a solution to such problems, and this will be supported with mathematically rigorous methods.

Problems of shortest paths are usually treated as discrete problems in graph theory. Practically, this approach is good enough but some conventional algorithms (A* algorithm, Dijkstra's algorithm) for instance do not guarantee that the chosen path is indeed the shortest.[1]

Because of this, the problem will be treated purely analytically so that different, hopefully more effective methods for solving the problem at hand, can be utilized. Due to the lack of literature concerning such approach, suitable definitions and notations involved in the problem will firstly be introduced, namely the idea of an obstacle and an admissible curve. From an analytical perspective, the problem can be further classified as a constrained variational problem.

2 Theoretical part

In order to develop an elementary understanding of the problem, calculus of variations along with theorems of real analysis will be studied. The concept of a functional and first variation will be introduced while omitting details that can be found in books on calculus of variations and real analysis.[2,3]

2.1 Calculus of variations

Many actions in nature can be quantified with an appropriate functional, usually given in an integral form. For example, a path described by a smooth function $y(x)$ on an interval $[a, b]$ is measured by $\int_a^b \sqrt{1+(y')^2}dx$. This can intuitively be shown by slicing up the interval into small pieces and measuring the length of each small hypotenuse described by by the change in x, dx and the change in y, dy.

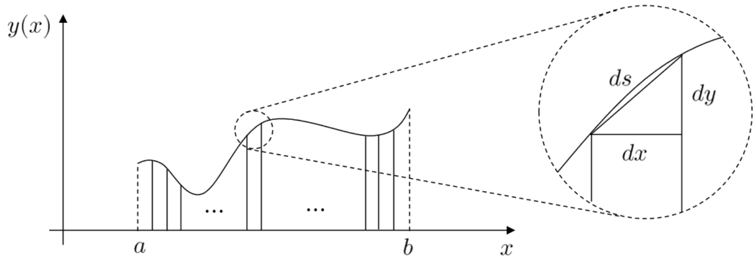

Figure 1: Understanding the length functional

The infinitesimal lengths ds can then be added up along the interval $[a, b]$ by taking the integral over $[a, b]$. Since dx goes to 0, for the derivative of $y(x)$ the equality $y' = \frac{dy}{dx}$ holds. Therefore:

$$L(y) = \int_a^b ds = \int_a^b \sqrt{dx^2 + dy^2} = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int_a^b \sqrt{1 + (y')^2} dx$$

Definition 2.1. *Suppose a function y has a continuous derivative on $[a, b]$. The length of y on $[a, b]$ is defined by the integral*

$$L(y) = \int_a^b \sqrt{1 + (y')^2} dx \tag{1}$$

Definition 2.2. *A functional is a function*

$$I : \mathcal{F} \ni y \to I(y) \in \mathbb{R},$$

where $\mathcal{F}$ is a space or a set of functions.

The natural question is then to ask what function minimizes the functional $L(y)$. Calculus of variations provides a necessary condition for an extreme of a functional in form of the first variance and the Euler-Lagrange differential equation.

Theorem 2.1. *If a function y is a local extreme of the functional $I(y)$ that is defined on $\mathcal{F}$ and h is a function such that $y + th \in \mathcal{F}$ for all small $t \in (-\varepsilon, \varepsilon)$, then*

$$\delta I(y)h = \lim_{t \to 0} \frac{I(y + th) - I(y)}{t} = 0.$$

The expression $\delta I(y)h$ is called the first variance of the functional I in y in the direction of h.

Theorem 2.2. *Suppose a functional I is of the form $I(y) = \int_a^b F(x, y, y')dx$, where F is continuously partially differentiable with respect to all there variables and y is a smooth function of x. If y is a local extreme of I and has fixed endpoints: $y(a) = A, y(b) = B$, then it must satisfy the Euler-Lagrange equation:*

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0. \tag{2}$$

This is one of the most important theorems of calculus of variations. Proof will be omitted but can be found in books on the topic.[2] The equation can be applied to the length functional (1) to find the shortest path between the points $[a, A]$ and $[b, B]$. By taking the appropriate partial derivatives, the equation (2) becomes:

$$0 - \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}}\right) = 0.$$

The solution to this differential equation is of the form $y(x) = c_1 x + c_0$ which with $c_1 = \frac{B-A}{b-a}$ and $c_0 = B - \frac{B-A}{b-a}$ is a straight line going through the desired points. With a little more effort, it can be shown that this solution is the global minimizer. Therefore, the shortest path between two points in a plane is a straight line. Many other functional can be formed to describe different aspects of a curve. For example, the functional $T(y) = \int_a^b \frac{\sqrt{1+(y')^2}}{\sqrt{y}} dx$ gives a total time in which a particle gets from $[a, A]$ to $[b, B]$ in a homogeneous gravitational field. The solution of the corresponding Euler-Lagrange equation then gives the path of shortest time - The Brachistochrone. Other examples along with solutions and with proofs of preceding theorems can be found in books on calculus of variations of authors paper.[2,4]

2.2 Theorems of real analysis

In order to support the intuition behind the original problem with mathematically rigorous methods, some of the most important theorems of real analysis will be utilised. Their proofs and other useful theorems can be found in books on mathematical analysis.[3]

Theorem 2.3 (Bolzano's theorem). *If a real function f is continuous on $[a, b]$ and satisfies $f(a) \cdot f(b) \leq 0$, then there exists $c \in [a, b]$ such that $f(c) = 0$.*

Theorem 2.4 (Mean Value Theorem). *If a function f is differentiable on $[a, b]$, then there exists $c \in [a, b]$ such that $f'(c) = \dfrac{f(b) - f(a)}{b - a}$.*

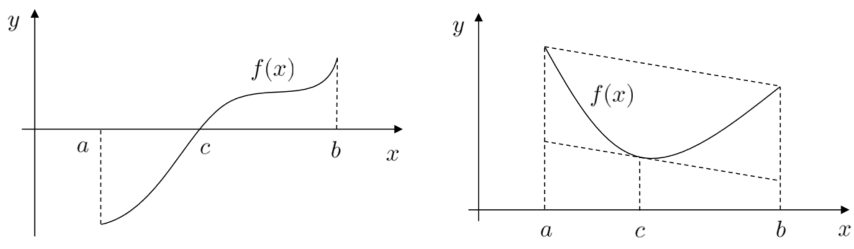

Figure 2: Bolzano's Theorem and Mean Value Theorem

These two theorems will be used as lemmas when proving that a given curve is indeed the shortest one. Since convexity and concavity play a role in the problem, definition of convex set and concave function will now be given.

Definition 2.3. *Suppose X is a subset of a vector space E. The set X is called convex if $\forall x, y \in X \wedge t \in [0, 1]$ the line segment $tx + (1 - t)y \in X$.*

Definition 2.4. *A real function f is said to be concave on an interval X, if $\forall x_1, x_2 \in X \wedge \forall t \in [0, 1]$ the inequality*

$$f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2)$$

holds.

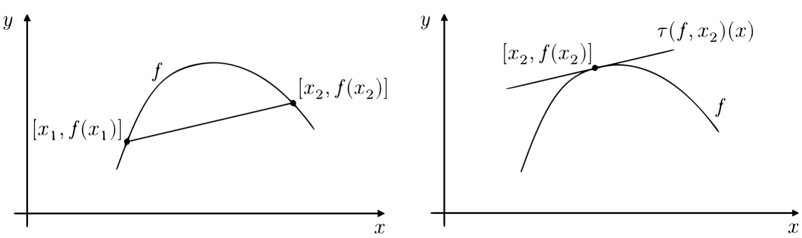

Figure 3: Properties of real concave functions

From this definition it can additionally be shown that the tangents to a concave function lie above the function, the first derivative of a concave function is not increasing and its second derivative is non-positive.[5]

3 Results

Firstly, we will provide definitions of obstacles and admissible curves. It will later be shown that other useful properties can be derived from these definitions.

Definition 3.1. *An obstacle $\mathcal{P} \in \mathbb{R}^2$ is given by a pair $(p_1, >)$ or $(p_2, <)$, where p_1, p_2 are continuous functions. The obstacles themselves are defined as the following sets:*

$$\mathcal{P}(p_1, >) = \{[a_x, a_y] | a_y > p_1(a_x)\}$$

$$\mathcal{P}(p_2, <) = \{[b_x, b_y] | b_y < p_1(b_x)\}$$

The functions describing the obstacle may have a restricted range to a closed interval.

Definition 3.2. *Suppose desired endpoints $[a, A], [b, B]$ are given along with obstacles $\mathcal{P}_1, ..., \mathcal{P}_n$ on $[a, b]$. A differentiable curve y is said to be admissible if:*

$$\forall x \in [a, b], \forall m = 1, ..., n \,:\, [x, y(x)] \notin \mathcal{P}_m \wedge y(a) = A, y(b) = B. \tag{3}$$

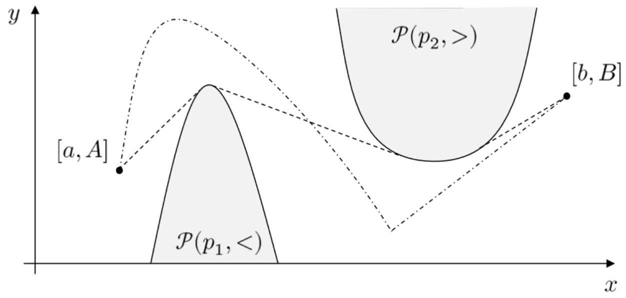

Figure 4: Visualisation of the problem

3.1 One obstacle with fixed endpoints

Suppose only one obstacle $\mathcal{P}(p, <)$ described by smooth concave function p is given along with two endpoints that lie on the obstacle: $p(a) = A, p(b) = B$. The set of all admissible curves $\mathcal{A}$ is of the form:

$$\mathcal{A} = \{y | y \text{ is differentiable}, y(x) \geq p(x) \forall x \in [a, b], y(a) = p(a), y(b) = y(p)\}. \tag{4}$$

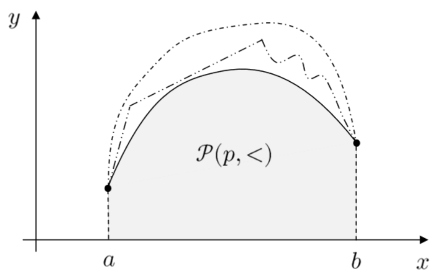

Figure 5: Problem with a concave function

Intuitively, the shortest path leads along the obstacle, that is, the curve $y = p$ is the shortest one.

Theorem 3.1. *Suppose we are given an obstacle $\mathcal{P}(p, <)$, where p is differentiable, and two endpoints $A = p(a), B = p(b)$. Under the assumption of an existence of a global minimizer, the path $y = p$ is the shortest one in $\mathcal{A}$.*

Proof. We proceed by contradiction. Suppose that shortest path satisfies $y(c) > p(c)$ for some $c \in (a, b)$. In this case we can define a new function $y_{pw}(y)$ the following way:

$$y_{pw}(x) = \begin{cases} y(x) & \text{for } x \in [a, e_a] \\ p'(c)(x - c) + p(c) & \text{for } x \in [e_a, e_b] \\ y(x) & \text{for } x \in [e_b, b] \end{cases}$$

, where e_a, e_b are points such that $p'(c)(e_a - c) + p(c) = y(e_a)$ and $p'(c)(e_b - c) + p(c) = y(e_b)$. Their existence is guaranteed by Bolzano's theorem 2.3 and the concavity of p. The length of the original curve y can be broken up into three integrals:

$$L(y) = \int_a^{e_a} \sqrt{1 + (y')^2} dx + \int_{e_a}^{e_b} \sqrt{1 + (y')^2} dx + \int_{e_b}^{b} \sqrt{1 + (y')^2} dx.$$

The length of y_{pw} can be evaluated from its definition:

$$L(y_{pw}) = \int_a^{e_a} \sqrt{1 + (y')^2} dx + \int_{e_a}^{e_b} \sqrt{1 + (p'(c))^2} dx + \int_{e_b}^{b} \sqrt{1 + (y')^2} dx.$$

Because $p'(c)(x - c) + p(c)$ is a straight line between $[e_a, y(e_b)]$ and $[e_b, y(e_b)]$ and y is not straight on $[e_a, e_b]$ by definition, we have $L(y_{pw}) < L(y)$, which is a contradiction, because y was assumed to be the shortest path. Therefore, if there is a minimizer, it must be p. This completes the proof. □

In this basic example, one can proceed without needing to assume the existence of a minimizer. Traditionally, direct methods of calculus of variations are used to prove statements of such nature. However, the following theorem is fairly elementary and does not require any background in these methods. On the other hand, it works with a stronger assumption on p.

Theorem 3.2. *Suppose we are given an obstacle $\mathcal{P}(p, <)$, where p is twice differentiable and two endpoints $A = p(a), B = p(b)$. The global minimizer is the path $y = p$.*

Proof. We wish to prove that for a concave function p on $[a, b]$ and a differentiable function $d(x) \geq 0$ on $[a, b]$ and $d(a) = d(b) = 0$, the inequality

$$\int_a^b \sqrt{1 + (p' + d')^2} dx \geq \int_a^b \sqrt{1 + (p')^2} dx$$

holds. Evidently, $(d')^2 \geq 0$ for any real function on $[a, b] \subset \mathbb{R}$. This is equivalent to

$$[1 + (p' + d')^2][1 + (p')^2] \geq [p'd' + 1 + (p')^2]^2$$

for any p'. Since both sides of the inequality are positive, square root on both sides can be taken and the inequality sign kept:

$$\sqrt{1 + (p' + d')^2}\sqrt{1 + (p')^2} \geq p'd' + 1 + (p')^2$$

After dividing both sides by $\sqrt{1 + (p')^2}$ and rearranging, the inequality becomes

$$\sqrt{1 + (p' + d')^2} - \sqrt{1 + (p')^2} \geq \frac{p'd'}{\sqrt{1 + (p')^2}} \tag{5}$$

By integrating both sides on $[a, b]$ we get the following integral inequality:

$$\int_a^b \sqrt{1 + (p' + d')^2} dx - \int_a^b \sqrt{1 + (p')^2} dx \geq \int_a^b \frac{p'd'}{\sqrt{1 + (p')^2}} dx.$$

On the right side we can proceed with integration by parts. Let

$$u = \frac{p'}{\sqrt{1+(p')^2}} \qquad v' = d'$$

$$u' = \frac{p''\sqrt{1+(p')^2} - \frac{(p')^2}{\sqrt{1+(p')^2}}}{1+(p')^2} \qquad v = d$$

$$\int_a^b \frac{p'd'}{\sqrt{1 + (p')^2}} dx = \left[\frac{p'}{\sqrt{1 + (p')^2}} \cdot d\right]\Bigg|_a^b + \int_a^b \left[\frac{\frac{(p')^2}{\sqrt{1+(p')^2}} - p''\sqrt{1 + (p')^2}}{1 + (p')^2} \cdot d\right] dx$$

The first summand on the right side vanishes since $d(a) = d(b) = 0$. Since p is concave, then $-p'' \geq 0$. This means that the whole integrand is positive, therefore, the whole integral must be positive, finally giving us the desired inequality:

$$\int_a^b \sqrt{1+(p'+d')^2}dx - \int_a^b \sqrt{1+(p')^2}dx \geq \int_a^b \frac{p'd'}{\sqrt{1+(p')^2}}dx \geq 0$$

$$\int_a^b \sqrt{1+(p'+d')^2}dx \geq \int_a^b \sqrt{1+(p')^2}.dx$$

□

3.2 Convexity of $\mathcal{A}$ and the functional L

Proving that a given set or a function is convex is useful, because one can consequently resort to more methods of solving a given problem using already developed methods of convex optimization.[5] In this chapter, we will demonstrate the convexity of the set $\mathcal{A}$ defined by (3) and the length functional $\int_a^b \sqrt{1+(y')^2}dx$.

Theorem 3.3. *Let $C^1[a,b]$ be the set of all differentiable functions on $[a,b]$. The subset $\mathcal{A} \subset C^1[a,b]$ defined by (3) is a convex set.*

Proof. Without loss of generality, we will only prove that the set (4) is convex. Let $y_1, y_2 \in \mathcal{A}$. Then the function

$$y(x) = ty_1(x) + (1-t)y_2(x),$$

where $x \in [a,b] \wedge t \in [0,1]$. y is differentiable, since y_1, y_2 are differentiable and t is a constant with respect to x. Since $t \geq 0$ and $1-t \geq 0$ the two inequalities $ty_1(x) \geq tp(x), (1-t)y_2(x) \geq (1-t)p(x)$ can be added to obtain

$$y(x) = ty_1(x) + (1-t)y_2(x) \geq tp(x) + (1-t)p(x) = p(x).$$

Also, $y_1(a) = y_2(a) = p(a)$ and $y_1(b) = y_2(b) = p(b)$ and therefore

$$y(a) = ty_1(a) + (1-t)y_2(a) = tp(a) + p(a) - tp(a) = p(a),$$

$$y(b) = ty_1(b) + (1-t)y_2(b) = tp(b) + p(b) - tp(b) = p(b).$$

The function satisfies all three conditions in (4) and therefore $y(x) = ty_1(x)+(1-t)y_2(x) \in \mathcal{A} \ \ \forall y_1, y_2 \in \mathcal{A} \wedge t \in [0,1]$, making the set $\mathcal{A}$ convex. □

Theorem 3.4. *The length functional $I(y) = \int_a^b F(x,y,y')dx$ is convex over all curves $\mathcal{A} \subset C^1[a,b]$.*

Proof. The functional $I(y) = \int_a^b F(x,y,y')dx$ is said to be convex, if it satisfies

$$F(x, y+h, y'+k) - F(x,y,y') \geq h\frac{\partial F}{\partial y} + k\frac{\partial F}{\partial y'} \tag{6}$$

$\forall [x,y,y',h,k] \in [a,b] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.[2] For the integrand $F = \sqrt{1+(y')^2}$, where $\frac{\partial F}{\partial y} = 0$ and $\frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{1+(y')^2}}$ with $[h,k] = [d,d']$, the inequality (6) is equivalent to (5), making the length functional $L(y) = \int_a^b \sqrt{1+(y')^2}dx$ convex. □

3.3 One obstacle with free endpoints

Let us have an obstacle $\mathcal{P}(p,<)$, where p is a concave, continuously differentiable on $[\alpha,\beta] \subset [a,b]$ and two endpoints $[a,A],[b,B]$. If the straight line $l_{AB}(x)$ between the two endpoints is admissible, then it is also the shortest path. If it is not, we suppose that l_{AB} and p intersect at points α,β such that $a < \alpha < \beta < b$.

Theorem 3.5. *Under the above mentioned assumptions, there exist two points c_a, c_b such that the curve $y \in \mathcal{A}$ of the form*

$$y(x) = \begin{cases} p'(c_a)(x-c_a)+p(c_a) & \text{for } x \in [a,c_a] \\ p(x) & \text{for } x \in [c_a,c_b] \\ p'(c_b)(x-c_b)+p(c_b) & \text{for } x \in [c_b,b] \end{cases},$$

where $y(a) = A, y(b) = B$ exists and is the shortest one in $\mathcal{A}$.

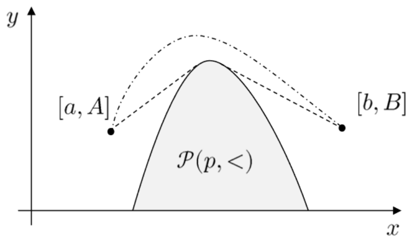

Figure 6: Second case of the problem

Proof. The existence of c_a and c_b is proven by the following argument. Without loss of generality, we only give proof of c_a's existence. If α and β exist, then by the Mean Value Theorem 2.4 there exists $c \in [\alpha,\beta]$ such that $p'(c) = \frac{p(\beta)-p(\alpha)}{\beta-\alpha} = \frac{B-A}{b-a}$, meaning that the tangent $p'(c)(x-c)+p(c)$ is parallel with l_{AB} and lies on or above l_{AB}, therefore $p'(c)(a-c)+p(a) \geq A$. The function p is concave, therefore p' is non-increasing, which means $p'(\alpha) \geq p'(c) = \frac{B-A}{b-a}$. Also, since $a-\alpha < 0$, the following holds:

$$p'(\alpha)(a-\alpha) \leq \frac{B-A}{b-a}(a-\alpha).$$

After adding $p(\alpha) = l_{AB}(\alpha) = \frac{B-A}{b-a}(\alpha-a)+A$ to both sides if the inequality, we yield the property

$$p'(\alpha)(a-\alpha)+p(\alpha) \leq A.$$

By defining a function

$$f(t) = p'(t)(a-t)+(t)$$

for $t \in [\alpha,c]$ and observing that $f(\alpha) \leq A, f(c) \geq A$, Bolzano's theorem 2.3 guarantees an existence of precisely c_a such that $p'(c_a)(a-c_a)+p(c_a) = A$. The existence of c_b can be proven analogously.

To show that $y(x)$ is the shortest path, one can proceed similarly as in the first case of the problem. Assume that the shortest path differs from y at least at one point on $[a,b]$ and construct a new path that is shorter than the path that was assumed to be the shortest one, arriving at a contradiction. Detailed proofs and other cases of the problem can be found in authors paper.[4] □

4 Discussion

This article treats a problem of finding an optimal path uniquely using the methods of mathematical analysis, more precisely, the methods of calculus of variations. Although the article does not provide a definite algorithm for finding the shortest path, it deals with more fundamental and theoretical aspects of the problem such as the length itself and its properties. The very existence of a solution is proven in one of the cases. This method might be extendable to prove the existence of a solution to any case of the problem. Even though the introduced definitions were constructed intuitively, they still had useful mathematical properties. Ideally, in the future research a necessary condition for the shortest path will be discussed and provided. This condition might be in form of a differential equation, similarly to the Euler-Lagrange equation.

5 Conclusion

Basic introduction from an analytical perspective was given along with definitions of basic concepts of the problem, namely the concept of length functional, set of all admissible curves and obstacles. The convexity of the set of all admissible curves and the length functional was demonstrated, which allowed us to prove the existence of a solution in one of the cases. Other case of the problem was considered, where we established under which conditions does a shortest path attain its general form and what this form looks like.

Acknowledgement

Most notably, I would like to thank professor Dušan Bednařík for introducing me to calculus of variations, for spending his free time discussing mathematics with me, providing literature and pushing me forward in my research.

References

(1) Cook, W., *In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation*; Princeton University Press: 2011, ISBN: 9781400839599.

(2) Coleman, R., *Calculus on Normed Vector Spaces*; Universitext; Springer New York: 2012, ISBN: 9781461438946.

(3) Rudin, W., *Principles of Mathematical Analysis*; International series in pure and applied mathematics; McGraw-Hill: 1976, ISBN: 9780070856134.

(4) Kloud, V., *Optimalizace délky křivek v rovině*; Students' Professional Activities: Hradec Králové, 2019, http://soc.nidv.cz/archiv/rocnik41/obor/1.

(5) Boyd, S.; Boyd, S.; Vandenberghe, L.; Press, C. U., *Convex Optimization*; Berichte über verteilte messysteme díl 1; Cambridge University Press: 2004, ISBN: 9780521833783.