# Fritz John's Equation in Mechanism Design

Alfred Galichon

# FRITZ JOHN'S EQUATION IN MECHANISM DESIGN

ALFRED GALICHON[§]

ABSTRACT. We show the role that an important equation first studied by Fritz John plays in mechanism design.

**Dedicated to Nicholas Yannelis on his 65th birthday.**

*Keywords:* implementability, mechanism design, John's equation, Kevin Roberts' theorem

*Date*: October 5, 2020. Funding from NSF grant DMS-1716489, as well as ERC grant CoG-866274 are acknowledged. The author benefited from insightful discussions with Guillaume Carlier, Deniz Dizdar, and Benny Moldovanu.

[§]New York University, departments of economics and mathematics and Sciences Po, department of economics; ag133@nyu.edu.

A large part of the literature on mechanism design deals with implementability in dominant strategy. Let us recall the basic result in the single-agent case, following Rochet (1987) and McAfee and McMillan (1988), and as exposited in Chapter 4.4 of Vohra (2011). Assume $x \in \mathbb{R}^d$ is the type reported by the agent, and $z \in \mathbb{R}^d$ is the outcome selected by the mechanism. The mechanism specified an allocation rule $T : \mathbb{R}^d \to \mathbb{R}^d$ and a payment rule $\pi : \mathbb{R}^d \to \mathbb{R}$. If the agent announces $x$, the outcome $z = T(x)$ is selected, while the agent is asked to pay $\pi(x)$. It is assumed that if the agent is of type $x$, if outcome $z$ is selected, and if the payment is $\pi$, the agent's utility is $x^\top z - \pi$. The mechanism is called implementable in dominant strategy (or simply implementable) if reporting her true type is the agent's dominant strategy; an allocation rule $T$ is called implementable in dominant strategy if there exists a payment rule $\pi$ such that the mechanism $(T, \pi)$ is implementable. This happens if

$$x^\top T(x) - \pi(x) \geq x^\top T\left(x'\right) - \pi\left(x'\right) \;\; \forall x' \in \mathbb{R}^d.$$

Denoting $V(x) = \max_{x' \in \mathbb{R}^d} \left\{x^\top T(x') - \pi(x')\right\}$, this will be the case when $T(x)$ is in the subdifferential of $V(x)$, or when $T$ is continuous, when $T(x) = \nabla u(x)$.

Hence the following result due to Rochet (1987) and McAfee and McMillan (1988):

**Theorem** (Implementation theorem)**.** *In the single-agent case, a continuous allocation rule $T : \mathbb{R}^d \to \mathbb{R}^d$ is implementable in dominant strategy if and only if $T(x) = \nabla V(x)$ for some convex function $V$.*

The purpose of this note is to investigate the multi-agent case. Assume that the space of types of each agent is still $\mathbb{R}^d$, and denote $x \in \mathbb{R}^d$ the type of the first agent and $y \in \mathbb{R}^d$ the type of the second agent. The outcome $z$ is still an element of $\mathbb{R}^d$, and the allocation rule is now a map $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, where $z = T(x, y)$ is the outcome selected if agent 1 announces type $x$ and agent 2 announces type $y$. The payment by agent 1 is $\pi_1(x, y)$ while the payment by agent 2 is $\pi_2(x, y)$. Denoting $V_1(x, y) = \max_{x' \in \mathbb{R}^d} \left\{x^\top T(x', y) - \pi_1(x', y)\right\}$ and $V_2(x, y) = \max_{y' \in \mathbb{R}^d} \left\{y^\top T(x, y') - \pi_2(x, y')\right\}$, it is easy to adapt the previous theorem to show that in the two-agent case, a continuous allocation rule $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is implementable in dominant strategy if and only if $T(x, y) = \nabla_x V_1(x, y)$ for some function

$V_1(x,y)$ which is convex in $x$ for all $y$, and $T(x,y) = \nabla_y V_2(x,y)$ for some function $V_2(x,y)$ which is convex in $y$ for all $x$.

The main result in this note is the following statement:

**Proposition.** *Consider a smooth allocation rule $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, and assume it is implementable. Then $T(x,y) = \nabla_x V_1(x,y)$ where $V_1$ satisfies* Fritz John's equation

$$\frac{\partial^2 V_1(x,y)}{\partial x_i \partial y_j} = \frac{\partial^2 V_1(x,y)}{\partial x_j \partial y_i}, 1 \leq i,j \leq d \tag{1}$$

*and in addition, the resulting symmetric matrix is semidefinite positive. Similarly, $T(x,y) = \nabla_y V_2(x,y)$ where $V_2$ satisfies the same restrictions.*

*Proof.* If $T$ is implementable, then $T(x,y) = \nabla_x V_1(x,y)$, where $V_1(x,y)$ is convex in $x$ for all $y$ and $T(x,y) = \nabla_y V_2(x,y)$ where $V_2(x,y)$ is convex in $y$ for all $x$. Because $T^i(x,y) = \partial V_2(x,y)/\partial y_i$, one has $\partial T^i(x,y)/\partial y_j = \partial^2 V_2(x,y)/\partial y_i \partial y_j$, and hence $\left(\partial T^i(x,y)/\partial y_j\right)_{ij}$ is symmetric semi-definite positive. But because $T$ is also a gradient with respect ot $x$, one has $T^i(x,y) = \partial V_1(x,y)/\partial x_i$, and thus

$$\frac{\partial^2 V_1}{\partial x_i \partial y_j}(x,y) = \frac{\partial T^i}{\partial y_j}(x,y) = \frac{\partial^2 V_2}{\partial y_i \partial y_j}(x,y),$$

which shows that $\left(\partial^2 V_1(x,y)/\partial x_i \partial y_j\right)_{ij}$ is symmetric semi-definite positive. Similarly, it is easy to see that

$$\frac{\partial^2 V_2}{\partial x_i \partial y_j}(x,y) = \frac{\partial T^j}{\partial x_i}(x,y) = \frac{\partial^2 V_1}{\partial x_i \partial x_j}(x,y),$$

and therefore $\left(\partial^2 V_2(x,y)/\partial x_i \partial y_j\right)_{ij}$ is also symmetric semi-definite positive. ∎

Equation (1) is a well-known mathematical equation appearing in harmonic analysis and inverse problems: it is called *Fritz John's ultrahyperbolic equation*, see John (1938), Kurusa (1991) and Ehrenpreis (2003). It plays an important role in medical imagery because of its connection with the so-called X-ray transform, a variant of the Radon transform; however, to the best of the author's knowledge, its occurrence in mechanism design problems seems to have remained unnoticed until now. Fritz John (1938) for $d = 3$, and Kurusa (1991) more generally provided rigorous conditions under which the solutions to (1) are given exactly by functions of the form

$$V_1(x,y) = \int_{-\infty}^{+\infty} \frac{1}{\lambda} \phi_\lambda (\lambda x + (1-\lambda) y)\, d\lambda \tag{2}$$

where $\phi_\lambda : \mathbb{R}^d \to \mathbb{R}$. Indeed,

$$\frac{\partial^2 \phi_\lambda ((1-\lambda) x + \lambda y)}{\partial x_i \partial y_j} = \lambda (1-\lambda) \frac{\partial^2 \phi_\lambda}{\partial w_i \partial w_j} ((1-\lambda) x + \lambda y)$$

is symmetric, and thus the sum is.

Note, however that while functions of the form (2) satisfy John's equation (1), they do not necessarily satisfy the positive semidefiniteness restriction that are expressed in the proposition. In order to ensure this restriction is satisfied, it is natural to restrict to $\lambda \in [0,1]$ and $\phi_\lambda$ convex, and thus introduce the class of solutions

$$V_1(x,y) = \int_0^1 \frac{1}{\lambda} \phi_\lambda (\lambda x + (1-\lambda) y)\, d\lambda$$

where $\phi_\lambda : \mathbb{R}^d \to \mathbb{R}$ are convex functions. This yields solutions of the form

$$T(x,y) = \int_0^1 T_\lambda (x,y)\, d\lambda,$$

where

$$T_\lambda (x,y) := \nabla \phi_\lambda (\lambda x + (1-\lambda) y),$$

and $T_\lambda (x,y)$ is called an *elementary allocation rule.*

Let us study the elementary allocation rules $T_\lambda (x,y)$. One has

$$\nabla \phi_\lambda (w) = \arg\max_{z \in \mathbb{R}^d} \left\{ w^\top z - \phi_\lambda^* (z) \right\},$$

where $\phi_\lambda^* (z) = \max_{w \in \mathbb{R}^d} \left\{ w^\top z - \phi_\lambda (w) \right\}$ can be interpreted as a payment rule. Hence,

$$T_\lambda (x,y) = \arg\max_{z \in \mathbb{R}^d} \left\{ \lambda x^\top z + (1-\lambda) y^\top z - \phi_\lambda^* (z) \right\}.$$

Note that $\lambda x^\top z + (1-\lambda) y^\top z$ is a measure of the social welfare where one assigns weight $\lambda$ to agent 1, and weight $(1-\lambda)$ to agent 2. Therefore, $T_\lambda (x,y)$ is an affine welfare maximizer. Note that when one imposes further that the set of outcomes should be finite and when

$d \geq 2$, a theorem by Kevin Roberts (1979) asserts that the only possible allocation rule should be the affine welfare maximizers [1]. Removing the restriction that the set of outcomes should be finite yields many more solutions – in particular, sums of affine welfare maximizers. A problem that seems interesting is to determine if when $d \geq 2$, there are implementable rules that are not affine welfare maximizers.

Let us take a very simple example:

**Example 1.** Consider a situation where two goods must be allocated between two players, so that each player gets one good. Player 1 has valuation $x_1$ for good 1 and $x_2$ for good 2, and player 2 has valuation $y_1$ for good 1, and $y_2$ for good 2. It is assumed that $x_1 > x_2$ and $y_1 < y_2$. Call "direct" the assignment where player 1 gets good 1 and player 2 gets good 2, and "reverse" the opposite assignment. Let $z_1$ be the probability of a direct assignment, and $z_2 = 1 - z_1$ the probability of a reverse assignment. The principal must decide on $z = (z_1, z_1)$ on the simplex. An implementable assignment rule is $z = T(x, y)$, where

$$T(x, y) = \left( \frac{x_1 - x_2}{x_1 - x_2 + y_2 - y_1}, \frac{y_2 - y_1}{x_1 - x_2 + y_2 - y_1} \right).$$

indeed, letting

$$V_1(x, y) = x_1 - (y_2 - y_1) \log(x_1 - x_2 + y_2 - y_1),$$

one verifies that $V_1(x, y)$ is convex in $x$, and that $T = \nabla_x V_1$, while letting

$$V_2(x, y) = y_2 - (x_1 - x_2) \log(x_1 - x_2 + y_2 - y_1)$$

one verifies that $V_2(x, y)$ is convex in $y$ and that $T = \nabla_y V_2$.

One has $\phi_\lambda(w) = \max\{w_1, w_2\}$ independent of $\lambda$, so that

$$\nabla \phi_\lambda(w) = (1\{w_1 \geq w_2\}, 1\{w_1 < w_2\}),$$

[1] Jehiel et al. (2008) study the notion of *cardinal potential* in the context of ex-post implementability, and derive a related partial differential equation which also bears a connection with Roberts' theorem, although they don't make the link with Fritz John's equation.

and when $w = \lambda x + (1-\lambda) y$, one has

$$T_\lambda (x,y) = \begin{pmatrix} 1\{\lambda (x_1 - x_2) + (1-\lambda)(y_1 - y_2) \geq 0\}, \\ 1\{\lambda (x_1 - x_2) + (1-\lambda)(y_1 - y_2) < 0\} \end{pmatrix}$$

and thus, integrating over $\lambda \in [0,1]$ with respect to the Lebesgue measure,

$$T(x,y) = \int_0^1 T_\lambda (x,y)\, d\lambda$$

that is

$$T(x,y) = \left( \frac{x_1 - x_2}{x_1 - x_2 + y_2 - y_1}, \frac{y_2 - y_1}{x_1 - x_2 + y_2 - y_1} \right).$$

This assignment rule can be interpreted as follows:

Draw $\lambda$ uniformly from $[0,1]$. Scale the valuation of player 1 by $\lambda$, and the valuations of player 2 by $(1-\lambda)$. Compute the valuation after rescaling associated with the direct and reverse assignment, respectively. Play the assignment which has whichever higher valuation.

## References


[1] Ehrenpreis, L. (2003). *The Universality of the Radon Transform*. Oxford University Press.

[2] Jehiel, P., Moldovanu, B., and Meyer-ter-Veh, M. (2008). "Ex-post Implementation and Preference Aggregation via Potentials". *Economic Theory* 37 (3), 469 – 490.

[3] John, F. (1938). "The ultrahyperbolic differential equation with four independent variables". *Duke Mathematical Journal* 4 (2), pp. 300–322.

[4] Kurusa, Á. (1991). "A characterization of the Radon transform's range by a system of PDEs." *Journal of Mathematical Analysis and Applications*, 161 (1), pp. 218–226.

[5] McAfee P. and McMillan, J. (1988). "Multidimensional incentive compatibility and mechanism design." *Journal of Economic Theory* 46 (2), pp. 335–354.

[6] Roberts, K. (1979). "The characterization of implementable choice rules". In Jean-Jacques Laffont, editor, *Aggregation and Revelation of Preferences*. Papers presented at the first European Summer Workshop of the Economic Society, pages 321–349. North-Holland.

[7] Rochet, J.-C. (1987). "A necessary and sufficient condition for rationalizability in a quasi-linear context." *Journal of Mathematical Economics* 16 (2), pp. 191–200.

[8] Vohra, R. (2011). *Mechanism Design: A Linear Programming Approach*. Cambridge University Press.