User Equilibrium versus System Optimum in Transportation when Costs are Non-separable and Asymmetric

Georgia Perakis**

*Sloan School of Management, MIT
77 Massachusetts Avenue, E53-359
Cambridge, MA 02139, USA
georgiap@mit.edu

1 Introduction

There has been an increasing literature in the recent years trying to quantify the inefficiency of Nash equilibrium problems (user-optimization) in non-cooperative games. The fact that there is not full efficiency in the system is well known both in the economics but also in the transportation literature (see [1]). This inefficiency of user-optimization was first quantified by Papadimitriou and Koutsoupias [11] in the context of a load balancing game. They coined the term "the price of anarchy" for characterizing the degree of efficiency loss. Subsequently, Roughgarden and Tardos [15] applied this idea to the classical network equilibrium problem in transportation with arc cost functions that are separable in terms of the arc flows. They established worst case bounds for measuring this inefficiency for affine separable cost functions and subsequently for special classes of separable nonlinear ones (such as polynomials). It should be noted that Marcotte presented in [12], results on the "price of anarchy" for a bilevel network design model. Recently, Johari and Tsitsiklis [9] also studied this problem in the context of resource allocation between users sharing a common resource. In their case the problem also reduces to one where each player has a separable payoff function. Correa, Schulz and Stier Moses [3] have also studied "the price of anarchy" in the context of transportation for capacitated networks. The cost functions they consider are also separable. The paper by Chau and Sim [2] has recently considered the case of nonseparable, symmetric cost functions giving rise to the same bound as Roughgarden and Tardos [15].

Wardrop [18] was perhaps the first to state the equilibrium principles in the context of transportation. Dafermos and Sparrow [4] coined the terms "user-optimized" and "system-optimized" in order to distinguish between Nash equilibrium where users act unilaterally in their own self interest versus when users are forced to select the routes that optimize the total network efficiency. Smith [16] and Dafermos [6] recognized that this problem can be cast as

*Acknowledgements: Preparation of this paper was partly supported, in part, by the PECASE Award DMI-9984339 from the National Science Foundation, and the Singapore MIT Alliance Program.

a variational inequality. In [5] Dafermos considered how the decentralized "user-optimized" problem can become a centralized "system optimization" problem through the imposition of tolls. Recently, Hearn and co-authors (see for example, [8]) have studied the problem of imposing tolls in order to induce a behavior to users so that their route choices are optimizing the overall system. They study a variety of criteria for imposing tolls. The review paper by Florian and Hearn [7], the book by Nagurney [13], and the references therein summarize the relevant literature in traffic equilibrium problems.

Nash equilibrium problems arise in a variety of settings and model competitive and non-cooperative behavior. In this paper we study the inefficiency of equilibrium by comparing how the presence of competition affects the total profit in the system in a decentralized (user-optimized) versus a centralized optimization (system-optimized) setting. We establish a bound on the ratio of the overall profit of the system in these two settings. This work is the first to consider non-separable, asymmetric cost functions and is important since it allows modeling more realistic situations. For example, the presence of congestion in a large transportation network where paths share arcs and there are several intersections, suggests that modeling these cost functions (often representing travel times) through separable functions (i.e., arc cost functions that depend only on the flow on that arc) may not be as realistic. In a large congested transportation network the travel time to traverse an arc will be affected by traffic congestion at its neighboring arcs. For example, the presence of a bottleneck or an accident at an arc ahead will slow traffic down at neighboring arcs as well. Furthermore, travel times are not affected by the flow on neighboring arcs in a symmetric way. For example, consider two consecutive arcs, then the travel time to traverse the first arc is not affected the same way by the flow of the arc ahead as the travel time of the arc ahead is affected by the flow of the arc behind. This discussion leads us to conclude that it is more realistic to consider non-separable, asymmetric cost functions in terms of the flow (see [10] for a discussion on how these travel times may be determined).

This work allows a unifying framework which naturally extends results in the current literature. In particular, our contributions versus the existing literature are the following.

1. We consider **non-separable** functions in the sense that cost functions also depend on the strategies of the competitors. Furthermore, cost functions can be **asymmetric** in the sense that different competitors' strategies affect their cost functions differently. This generalization is important since the strategies of one's competitors will influence his/her own cost in an asymmetric way. In particular, we introduce a measure of asymmetry (denoted by c^2 in Section 2) which quantifies **the degree of asymmetry** of the competitors' cost functions. We establish that the ratio of the total cost in the system operating in a user-optimized setting versus the total cost in a system optimized setting is bounded by

$$\begin{cases} \dfrac{4}{4-c^2} & \text{if } c^2 \leq 2 \\ c^2 & \text{if } c^2 > 2. \end{cases}$$

We illustrate how our results are a natural generalization of the bound becomes 4/3 as in [15] for separable problem functions and [2] for nonseparable symmetric ones. The results in the affine case allow the feasible region to be a non-convex set.

2. We generalize our results to **nonlinear** functions. We introduce a measure which quantifies **the degree of nonlinearity** of the problem function (denoted by A). We establish that the

bound naturally extends to involve the nonlinearity parameter A, i.e.,

$$\begin{cases} \dfrac{4}{4 - c^2 A} & \text{if } c^2 \leq \frac{2}{A} \\ c^2 A^2 - 2(A-1) & \text{if } c^2 > \frac{2}{A}. \end{cases}$$

3. We establish that the bound is **tight** for affine and for some nonlinear problems.
4. We introduce an alternative **semidefinite optimization** formulation for deriving these bounds. This approach does not require positive definiteness of the Jacobian matrix (i.e., it does not need to be invertible). Therefore, the solution does not need to be unique. We illustrate that this approach gives rise to the same bound when the Jacobian matrix is positive definite.

2 A Bound for Affine and Asymmetric Cost Functions

In this section, we establish a bound between the user and the system optimization problems in the context of minimizing cost. For the (UO) decentralized problem we will consider the variational inequality problem of finding $x_u \in K$ satisfying

$$F(x_u)^t(x - x_u) \geq 0, \quad \text{for all } x \in K. \tag{1}$$

Notice that in the traffic equilibrium context the variational inequality problem function F is the arc cost function vector while the vector of variables x are the arc flows. As we discussed in the previous section, it is more realistic to model this function as a non-separable asymmetric function of the flow.

Let x_u and x_s denote solutions of the user and system optimization problems respectively. Let $Z_u = F(x_u)^t x_u$ be the total cost for the user-optimized problem (UO) and $Z_s = F(x_s)^t x_s = \min_{x \in K} F(x)^t x$ be the total cost for the system-optimized problem (SO). In this section, we provide a bound on Z_u/Z_s for affine cost functions $F(x) = Gx + b$, with $G \succ 0$ (i.e., positive definite) and asymmetric matrix, $b^t x \geq 0$ for all $x \in K$ (notice that this follows when constant vector $b \geq 0$ and $K \subseteq R^n_+$ which is the case in the traffic equilibrium setting). In this case, the system optimization problem involves the minimization of a strictly convex quadratic function over the set K.

For a matrix G, we consider the symmetrized matrix

$$S = \frac{G + G^t}{2}$$

and introduce the following measure c^2 of the degree of asymmetry of matrix G:

Definition 1:

$$c^2 \equiv \|S^{-1}G\|_S^2 = \sup_{w \neq 0} \frac{\|S^{-1}Gw\|_S^2}{\|w\|_S^2} = \sup_{w \neq 0} \frac{w^t G^t S^{-1} G w}{w^t S w}.$$

Note that by setting $l = S^{1/2}w$, the previous definition of c^2 becomes

$$c^2 = \sup_{l \neq 0} \frac{l^t S^{-1/2} G^t S^{-1} G S^{-1/2} l}{\|l\|^2} = \lambda_{max}(S^{-1/2} G^t S^{-1} G S^{-1/2}).$$

When the matrix G is positive definite and symmetric, that is, $G = G^t$ (and therefore, $S = G$), then $c^2 = 1$. As an example, consider

$$G = \begin{bmatrix} 1 & a \\ -a & 1 \end{bmatrix}.$$

Since $S = I$, it easily follows that $c^2 = 1 + a^2$. The quantity c^2 in this case quantifies the observation that as $|a|$ increases, the degree of asymmetry of G increases as well.

Theorem 1 *(see [14])*
For an affine variational inequality problem with problem function $F(x) = Gx + b$, with $G \succ 0$, $b^t x \geq 0$ for all $x \in K$, we have:

$$\frac{Z_u}{Z_s} \leq \begin{cases} \dfrac{4}{4 - c^2} & \text{if } c^2 \leq 2 \\ c^2 & \text{if } c^2 > 2. \end{cases}$$

- **Separable affine cost functions:**
  When the variational inequality problem function F is separable, it has components $F_i(x) = g_i x_i + b_i$. In this case the matrix G is diagonal, with diagonal elements $g_i > 0$. In this case $c^2 = 1$ and the bound in Theorem 1 becomes

  $$\frac{Z_u}{Z_s} \leq \frac{4}{4 - c^2} = \frac{4}{3}$$

  originally obtained in Roughgarden and Tardos [15].

- **Non-separable symmetric affine cost functions:**
  When the variational inequality problem function F is non-separable, that is $F(x) = Gx + b$, with G a general symmetric positive definite matrix, then $c^2 = 1$ and thus $Z_u/Z_s \leq 4/3$, thus showing that the bound of $4/3$ holds also for non-separable symmetric affine functions (see also Chau and Sim [2]).

- **Non-separable asymmetric affine cost functions:**
  When the matrix G is "not too asymmetric" (in the sense that $c^2 \leq 2$) then the bound becomes $\frac{4}{4-c^2}$. On the other hand, for "rather asymmetric" matrices (in the sense that $c^2 > 2$) then the bound becomes c^2.

In [14] we establish that when the constant term is zero that is, $F(x) = Gx$, then the bound is always c^2. In [14] we illustrate that these bounds are tight.

3 A Bound for Nonlinear, Asymmetric Functions

In this section, we assume that the Jacobian matrix is not a constant matrix G but a positive definite, nonlinear and asymmetric matrix $\nabla F(x)$. The positive definiteness assumption of the

Jacobian matrix implies that the variational inequality problem has a unique solution (see for example [13] for details).

We introduce the symmetrized matrix, $S(x) = \frac{\nabla F(x)+\nabla F(x)^t}{2}$. We now extend Definition 1 for measuring the **degree of asymmetry** of the problem to a definition that also incorporates the nonlinearity of the cost functions involved.

Definition 2: We define a quantity c^2 that measures the degree of asymmetry of the Jacobian matrix $\nabla F(x)$. That is,

$$c^2 \equiv \sup_{x\in K} \|S(x)^{-1}\nabla F(x)\|^2_{S(x)}.$$

Constant c^2 is in this case the supremum over the feasible region, of the maximum eigenvalue of the positive definite and symmetric matrix

$$S(x)^{-1/2}\nabla F(x)^t S(x)^{-1}\nabla F(x) S(x)^{-1/2}.$$

When the Jacobian matrix is positive definite and symmetric, then $c^2 = 1$.

Furthermore, we need to define a measure of the **nonlinearity** of the problem function F. As a result, we consider a property of the Jacobian matrix which always applies to positive definite matrices. This allows us in some cases to provide a tight bound. This bound naturally extends the bound in Theorem 1 from affine to nonlinear problems. The bound involves the constant A that measure the nonlinearity of the problem.

Definition 3: (see [17] for more details)
The variational inequality problem function $F : R^n \to R^n$ satisfies **Jacobian similarity** property if it has a positive semidefinite Jacobian matrix ($\nabla F(x) \succeq 0,\ \forall x \in K$) and $\forall w \in R^n$, and $\forall x, \bar{x} \in K$, there exists $A \geq 1$ satisfying

$$\frac{1}{A} w^t \nabla F(x) w \leq w^t \nabla F(\bar{x}) w \leq A w^t \nabla F(x) w.$$

Lemma 1: (see [17]) The Jacobian similarity property holds under either of the following conditions:

- The Jacobian matrix is **strongly positive definite** (i.e. has eigenvalues bounded away from zero). Then a possible bound for the constant A is

$$A = \frac{\max_{x\in K} \lambda_{max}(S(x))}{\min_{x\in K} \lambda_{min}(S(x))}.$$

- The problem function is **affine**, with **positive semidefinite** Jacobian matrix G. In this case $A = 1$.

Theorem 2 *(see [14]) For a variational inequality problem with a strictly monotone, nonlinear continuously differentiable problem function F satisfying the Jacobian similarity property, $F(0)^t x \geq 0$ for all $x \in K$, we have:*

$$\frac{Z_u}{Z_s} \leq \begin{cases} \dfrac{4}{4 - c^2 A} & \text{if } c^2 \leq \frac{2}{A} \\ c^2 A^2 - 2(A-1) & \text{if } c^2 > \frac{2}{A}. \end{cases}$$

Remarks:

1. When the variational inequality problem function is affine, $F(x) = Gx+b$, then $\nabla F(x) = G$ and as a result $A = 1$ and the bound coincides with the one we found in the previous section.
2. When the term $F(0) = 0$ then the bound becomes $\frac{Z_u}{Z_s} \leq c^2 A^2 - 2(A-1)$.

References

[1] D. Braess, "Uber ein paradoxon der verkehrsplanung", *Unternehmenforschung*, 12, 258-268, 1968.

[2] C.K. Chau and K.M. Sim, "The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands", *Operations Research Letters*, 31, 2003.

[3] J. Correa, A. Schulz, and N. Stier Moses, "Selfish Routing in Capacitated Networks". MIT Sloan Working Paper No. 4319-03. (2003).

[4] S. Dafermos and F.T. Sparrow, "The traffic assignment problem for a general network", Journal of Research of the National Bureau of Standards - B, 73B, 1969.

[5] S. Dafermos, "Toll patterns for multiclass-user transportation networks", *Transportation Science*, 211-223, 1973.

[6] S. Dafermos, "Traffic equilibria and variational inequalities, *Transportation Science*, 14, 42-54, 1980.

[7] M. Florian and D. Hearn, "Network Equilibrium Models and Algorithms", *Handbook of Operations Research and Management Science* , vol. 8, (M. Ball, T. Magnanti, C. Monma and G. Neumhauser eds.), Elsevier Science B.V, 1995.

[8] D. Hearn and M. Ramana, "Solving congestion toll pricing models", *Equilibrium and Advanced Transportation Modeling*, editors P. Marcotte, S. Nguyen.

[9] R. Johari, and J.N. Tsitsiklis, "Network resource allocation and a congestion game", LIDS Working Paper, 2003.

[10] S. Kachani and G. Perakis, "Fluid Dynamics Models in Transportation and Pricing", Working Paper, ORC, MIT, 2002.

[11] E. Koutsoupias and C. H. Papadimitriou, "Worst-case equilibria," *STACS*, 1999.

[12] P. Marcotte, "Network design problem with congestion effects: A case of bilevel programming" Mathematical Programming 34 (1986) 142-162.

[13] A. Nagurney, "Network Economics; A Variational Inequality Approach" Kluwer Academic Publishers, 1993.

[14] G. Perakis, "The Price of Anarchy when Costs are Non-separable and Asymmetric", Working Paper, ORC, MIT, 2003.

[15] T. Roughgarden and E. Tardos, "Bounding the Inefficiency of Equilibria in Nonatomic Congestion Games", *Journal of the ACM*, 49(2):236–259, 2002.

[16] M.J. Smith, "The existence, uniqueness and stability of traffic equilibria", *Transportation Research*, 13B, 295 – 304, 1979.

[17] J. Sun, "A Convergence Analysis for a Convex Version of Dikin's Algorithm", *Annals of Operations Research*, 62, 357-374, 1996.

[18] J.G. Wardrop, "Some Theoretical Aspects of Road Traffic Research", in *Proceedings of the Institute of Civil Engineers*, Part II, 325-378, 1952.