RADAR

Research Archive and Digital Asset Repository

OXFORD BROOKES UNIVERSITY

Todinov, M

The same sign local effects principle and its application to technical risk reduction

Todinov, M (2015) The same sign local effects principle and its application to technical risk reduction. *International Journal of Reliability and Safety*, 9 (4).

doi: 10.1504/IJRS.2015.073137

This version is available: https://radar.brookes.ac.uk/radar/items/1e7917b9-2791-4101-a926-1edb0b704123/1/

Available on RADAR: July 2016

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the post print version of the journal article. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

WWW.BROOKES.AC.UK/GO/RADAR

The same sign local effects principle and its applications for risk reduction

M.T.Todinov

Department of Mechanical Engineering and Mathematical Sciences
Oxford Brookes University, Wheatley, Oxford OX33 1HX
E-mail: mtodinov@brookes.ac.uk

ABSTRACT

A simple yet powerful general risk-reduction principle has been formulated related to systems each state of which can be obtained from a given initial state by adding the effects from a specified set of modifications.

The proposed generic principle has also been applied with success to minimise the transportation costs related to a set of interchangeable sources servicing a set of destinations. A counter-example has been given which demonstrates that selecting the nearest available source to supply the destinations along the shortest available paths does not guarantee an optimal solution. This is contrary to some well-established practices in network optimisation. The application of the proposed generic principle in logistic supply networks leads to a significant reduction of the risk of congestion and delays. The associated reduction of transportation costs has the potential to save billions of dollars to the world economy.

Keywords: *generic principles, risk reduction, reliability improvement; optimization; multivariable functions; transportation network; logistic supply network*

To be cited: Todinov M.T., The same sign local effects principle and its application to technical risk reduction, *International Journal of Safety and Reliability* 9 (4) pp. 311-329, (2015).

1. INTRODUCTION

For a long time, principles for technical risk reduction have been exclusively oriented towards specific industries, technologies or operations, and usually have no general validity. Despite the critical importance of distilling generic principles related to reducing technical risk, very little has been published on this topic. French (1999) formulated a number of generic principles to be followed in conceptual design, but they were not necessarily oriented towards reducing technical risk. Collins (2003) and Pahl (2007) discussed engineering design with a failure prevention perspective. Most of the discussed principles however are either not related to reducing technical risk or are too specific with no general validity. Such is for example the principle of thermal design discussed in Pahl (2007). The formulated generic guidelines to be followed by engineer-designers were given in the specific context of mechanical design and no generic risk principles were formulated.

The available risk literature (Vose 2002; Aven 2003; Bedford & Cooke) which normally discusses more general aspects of technical risk reduction is oriented towards risk modelling, risk assessment, risk management and decision making and again, there is very little discussion on generic principles for reducing technical risk.

Reliability engineering literature usually focusses on predicting the reliability of components and systems rather than principles for reliability improvement. Some important generic principles for improving reliability like: *implementing active and standby redundancy*, *strengthening the weakest link*, *eliminating a common cause*, *reducing variability*, *reducing human errors* etc., have indeed been covered well in the reliability literature (Barlow & Proschan 1975; Ebeling, 1997; O'Connor, 2003; Lewis, 1996; Dhillon and Singh 1981). Their number however is not sufficient to cover the ever increasing demands of the design for reliability. A systematic formulation and classification of the generic principles for technical risk reduction was started in (Todinov 2007). The present work extends the number of existing generic principles for technical risk reduction by contributing a new generic principle, based on a general result referred to as '*the same sign local effects principle*'. To the best of our knowledge, this risk reduction principle has never been reported before.

The proposed principle is truly universal and can be applied not only in engineering but also in such diverse areas as logistics supply, management, economics, telecommunications, etc. The presented principle prompts design-engineers and risk managers not to limit themselves within few familiar ways of improving reliability and reducing risk often leading to solutions which are far from optimal.

Global flows of goods, services, people and finances have created a tightly interconnected world economy. The increasing global prosperity and growing pervasiveness of internet connectivity and digital technologies created a huge number of destinations for consumer demand and sources of production for goods and services. The past decade can be characterised by an explosion of data and communication flows. The global online traffic has grown from 84 petabytes (1 petabyte = 1 million gigabytes) a month in 2000 to more than 40000 petabytes a month in 2012 (McKinsey, 2014).

There is a strong correlation between the magnitude and reliability of flows of goods, services, data and finances and economic growth (McKinsey, 2014). A *connection* is a unique sequence of edges, lines, trajectories, etc. which provide the link between a source and a destination. Unreliable, expensive, congested or slow connections can wreak havoc on economic growth. The reliable and cost-efficient operation of the flows depends significantly on minimising the detrimental effects along the connecting paths.

In this paper, the same sign local effects principle will be applied to reduce risk on a network including interchangeable sources connected to a number of destinations.

2. THE SAME SIGN LOCAL EFFECTS PRINCIPLE

Consider a system in a state S_0 which is associated with a particular total effect E_0. The system can enter into any other possible state S_i. Suppose that the transition from state S_0 to any other possible state S_p can be done through a set Ω *of permitted local modifications* $\Delta e_i \in \Omega$, $i = 1,...,k$. A system with such property will be referred to as *system connected with the initial state, on the set* Ω *of permitted local modifications*.

Each local modification Δe_i belonging to the permitted set Ω, produces a local effect described by a certain function $f_i(\Delta e_i)$. The total effect associated with a particular system state S_p will be denoted by E_p.

Suppose that the total effect E_p associated with the possible system state p is a sum of the initial total effect E_0 associated with the initial state and the local effects $f_i(\Delta e_i)$ associated with the permitted local modifications Δe_i ($i = 1,...,k$):

$$E_p = E_0 + f_1(\Delta e_1) + f_2(\Delta e_2) + ... + f_k(\Delta e_k) \tag{1}$$

A system with such property will be referred to as *a system with additive response*.

The local effect $f_i(\Delta e_i)$ associated with any permitted local modification from the specified set can be nonnegative or negative.The following principle then holds.

The same sign local effects principle: *For a system with additive response, connected on a set of permitted local modifications, a sufficient condition for a global minimum, is the positive sign of all local effects bringing the system from the initial state to any possible system state.*

A sufficient condition for a global maximum is the negative sign of all local effects bringing the system from the initial state to any possible system state.

Proof.

Only the first half of the principle related to existence of a global minimum will be proved. The proof of the second half is analogous.

Because the system is connected with the initial state on the set Ω of specified modifications and is also a system with additive response, the total effect associated with any possible state S^* can always be obtained from the total effect of the initial state S_0 by adding a finite number k of local effects from k local modifications. Hence, the total effect E^* associated with state S^* is

$$E^* = E_0 + f_1(\Delta e_1) + f_2(\Delta e_2) + ... + f_k(\Delta e_k) \tag{2}$$

where $f_i(\Delta e_i)$ are the local effects resulting from the permitted local modifications Δe_i, on the defined set of modifications Ω.

According to the condition of the proposed principle, in transferring from state S_0 to any possible state S^*, there is no permitted local modification Δe_i whose local net effect is negative $f_i(\Delta e_i) < 0$. In other words, the possible local modifications bringing the system from state S_0 to a state S^*, result only in $f_i(\Delta e_i) > 0$. Then, the initial state S_0 is associated with the smallest total effect.

Indeed, suppose that a system state S^* exists, associated with a smaller total effect $E^* < E_0$ than the total effect associated with the initial state S_0. Because $E^* < E_0$ from equation (2), it follows that

$$f_1(\Delta e_1) + f_2(\Delta e_2) + ... + f_k(\Delta e_k) < 0 \tag{3}$$

must necessarily hold.

Consequently, there must be at least one permitted local modification, for which $f_i(\Delta e_i) < 0$. This means that one of the additive local effects is negative $f_i(\Delta e_i) < 0$. This however, contradicts the assumption that all permitted local modifications from the defined set Ω result only in positive local effects $f_i(\Delta e_i) > 0$ irrespective of the current levels of the input parameters. This completes the proof. □

The same sign local effects principle is a generic principle which provides the valuable opportunity *to quickly find a global extremum of a function by showing that all possible local modifications on a permitted set of modifications are of the same sign, which is often a considerably simpler task.* As a result, the stated principle provides an efficient strategy for guaranteeing a global extremum in many practical problems.

The proposed simple principle, similar to the pigeonhole principle in combinatorics (Brualdi, 2010), has deep and non-trivial applications. It often provides an efficient strategy for attaining a global extremum of a connected system with additive response through series of local modifications, which is often a simple task, achieved within a finite number of steps.

In what follows, this principle will be applied for (i) finding the global maximum/minimum of multivariable functions whose partial derivatives do not change sign in a specified domain and (ii) minimising the transportation cost on a network of connections from a number of interchangeable sources to the same number of destinations.

3. GLOBAL EXTREMUM OF MULTI-VARIATE FUNCTIONS WHOSE PARTIAL DERIVATIVES DO NOT CHANGE SING ON A SPECIFIED DOMAIN

The same sign local effects principle holds even if the effects from modifying the separate input variables depend on the current states of the rest of the input variables. Here is a simple illustrating example:

The function

$$f(x,y) = x^3 y^3 \tag{4}$$

is strictly increasing with respect to any of the input variables *x* and *y* within the domain: $0.1 \le x \le 10$; $0.1 \le y \le 10$. This is easy to verify because the first partial derivatives $\frac{\partial f}{\partial x} = 3x^2 y^3 > 0$ and $\frac{\partial f}{\partial x} = 3y^2 x^3 > 0$ are strictly increasing with respect to the input variables *x* and *y* in the specified function domain.

Consider the initial state $x = 0.1$ and $y = 0.1$. The set Ω of permissible local modifications include: (i) increasing the input variable *x* by a certain value and (ii) increasing the input variable *y* by a certain value. From the initial state ($x = 0.1$, $y = 0.1$) any possible other state x^*, y^* (point in the function domain) can be reached by increasing *x* by an appropriate value Δx followed by increasing *y* by an appropriate value Δy. The total effect associated with state x^*, y^* can be presented as

$$f(x^*, y^*) = f(x_0, y_0) + \Delta f_1 + \Delta f_2$$

where Δf_1 is the local effect due to increasing the input variable *x* by Δx and Δf_2 is the local effect due to increasing the input variable *y* by Δy. The sign of each of these local effects on the function are positive ($\Delta f_1 > 0, \Delta f_2 > 0$), therefore at $x = 0.1$ and $y = 0.1$ the function has a global minimum. The local effect Δf_2 from increasing the input variable *y* depends on the current value ($x_0 + \Delta x$) of the other input variable *x*. The local effect Δf_1 from increasing the input variable *x* also depends on the current value (y_0) of the other input variable *y*.

The positive sign of the local effects is guaranteed by the positive signs of the partial derivatives in the function domain.

Consider now the function

$$f(x,y) = -x^2 + 4x - y^3 - \sin y + \cos y - 4 \tag{5}$$

The partial derivative $\frac{\partial f}{\partial x} = -2x + 4$ of this function is with a positive sign in the domain $0.1 \le x \le \pi/2$; $0.1 \le y \le \pi/2$ while the partial derivative $\frac{\partial f}{\partial y} = -3y^2 - \cos y - \sin y$ is with a negative sign in the same domain.

Let us define the following set Ω of permissible local modifications: (i) increasing the input variable *x* by a certain value and (ii) decreasing the input variable *y* by a certain value.

Consider the initial state $x_0 = 0.1$ and $y_0 = \pi/2$. From this initial state any other state (function value) can be reached by increasing *x* by Δx and decreasing *y* by Δy.

In both cases the sign of the local effects on the function is positive, therefore at $x_0 = 0.1$ and $y_0 = \pi/2$ the function has a global minimum.

Now consider the initial state $x_1 = \pi/2$ and $y_1 = 0.1$, with the set Ω of permissible local modifications: (i) decreasing the input variable *x* by a certain value and (ii) increasing the input variable *y* by a certain value. From this initial state any other state (function value) can be reached by decreasing *x* by an appropriate value Δx and increasing *y* by an appropriate value Δy. The sign of each of the local effects on the function will be negative, therefore at $x_1 = \pi/2$ and $y_1 = 0.1$ the function will possess a global maximum.

These results can now be generalized for an arbitrary function $f(x_1, x_2, ..., x_n)$, where the input variables $x_1, x_2, ..., x_n$ vary within a particular rectangular domain:

$$a_1 \le x_1 \le b_1$$
$$a_2 \le x_2 \le b_2$$
$$.......$$
$$a_n \le x_n \le b_n$$

and whose partial derivatives do not change sign in the specified domain.
The coordinates of an initial state (point) are specified as follows. If the partial derivative with respect to x_i is positive in the domain, the left end a_i of the interval of variation of x_i is selected as the *i*th coordinate of the initial point. If the derivative with respect to x_i is negative in the domain, the right end b_i of the interval of variation of x_i is selected as the *i*th coordinate of the initial point. At the end of this process, for example, a point with coordinates $a_1, b_2, a_3, a_4, ..., b_n$ will be specified which defines the initial point. The as-specified initial point corresponds to a global minimum of the function in the specified domain.

Indeed, if the set of permissible modifications is specified as follows: increase the value of the variable x_i if it corresponds to a positive partial derivative and decrease the value of the variable x_i if it corresponds to a negative partial derivative, any possible point from the domain of the function $f(x_1, x_2, ..., x_n)$ can be reached by starting from the corner (point) with coordinates determined as described earlier.

The value y^* of the function at any point $x_1^*, x_2^*, ..., x_n^*$ is then obtained from adding a set of finite changes (increments):

$$y^* = y_0 + \Delta y_1 + \Delta y_2 + ... + \Delta y_n$$

where Δy_i is the local effect from changing by Δx_i the input variable x_i while the rest of the input variables are kept constant. During this process, the local effects Δy_i due to changing the input variables corresponding to a positive partial derivative will be incremented and will cause a positive increment Δy_i in the function value while variables corresponding to a negative partial derivative will be decreased and will also cause a positive increment Δy_i in the function value. Hence, according to the formulated principle, the as-defined initial state $a_1, b_2, a_3, a_4, ..., b_n$ corresponds to a global minimum of the function in its domain.

Each subsequent local effect Δy_i is dependent on the current values of the other inputs. Therefore, the local effects Δy_i depend on the current states of the rest of the input variables

Now consider an initial point whose coordinates are specified as follows. If the partial derivative with respect to x_i is positive, the right b_i end of the interval of variation of x_i is selected as the *i*th coordinate of the initial point. If the partial derivative with respect to x_i is negative, the left end a_i of the interval of variation of x_i is selected as the *i*th coordinate of the initial point. At the end of this process for example, a point with coordinates $a_1, b_2, a_3, a_4, ..., b_n$ will be defined as initial point. The as-specified initial point now corresponds to a global maximum. The justification is very similar to the one related to the global minimum and will not be repeated here.

As a result, an important application of the formulated principle has been found in determining the global extremum of multivariable functions whose partial derivatives maintain the same sign in a specified domain.

This application is closely related to improving the performance of heuristic methods searching for a global extremum of a multivariable function $f(x_1, x_2, x_3, ..., x_n)$ in a large domain. Many heuristic optimisation methods such as the *particle swarm optimisation* (Lazinica 2009, Kiranyaz et al., 2014) for example, incorporate a local search in the vicinity of a point to find the local best value. Because within a relatively small vicinity of a point the derivatives do not normally change sign, the proposed method can be used for determining quickly the extreme local value in the vicinity of a specified point. If the local search space is defined by *n*-dimensional rectangular domain

$$a_1 \le x_1 \le b_1;\ a_2 \le x_2 \le b_2; ....; a_n \le x_n \le b_n,$$

to determine the function values in each of the corner points of this domain involves 2^n evaluations of the function. This is an expensive and time consuming step if *n* is relatively large. With the proposed method, if the local search domain is relatively small, the signs of the partial derivatives can be checked just by obtaining the sign of the following finite differences:

$f(b_1, a_2, a_3, ..., a_n) - f(a_1, a_2, a_3, ..., a_n)$, $f(a_1, b_2, a_3, ..., a_n) - f(a_1, a_2, ..., a_n)$,

$f(a_1, a_2, b_3, ..., a_n) - f(a_1, a_2, a_3, ..., a_n)$,...., $f(a_1, a_2, a_3, ..., b_n) - f(a_1, a_2, a_3, ..., a_n)$

which involves only *n+1* evaluations of the function $f(x_1, x_2, x_3, ..., x_n)$. Depending on the signs of these finite differences, the corner corresponding to the local best value (maximum or a minimum) is determined immediately, according the method described earlier. One more final evaluation of the function is finally needed to find determine the value of the extremum. This results in only $n+2$ evaluations of the function $f(x_1, x_2, x_3, ..., x_n)$. As a result, the time complexity of this critical step is reduced from exponential $O(2^n)$ to linear $O(n)$ which results in a significantly improved performance of the optimisation algorithm.

4. APPLICATION OF THE PROPOSED PRINCIPLE TO A SYSTEM OF CONNECTING PATHS FROM INTERCHANGEABLE SOURCES TO DESTINATIONS

Apart from finding the global extremum of special classes of multivariable functions, there are other real-life systems that can be successfully optimized by using the proposed principle. Some of these involve interchangeable sources connected to a number of destinations. Here are some examples:

● *Transportation and delivery services from interchangeable supply centres:*

- Warehouses of the same company delivering commodity to various destinations
- Supermarkets from the same chain servicing clients
- Fuel terminals belonging to the same company delivering fuel to petrol stations
- Airline or bus companies servicing a number of destinations
- Mail delivery services
- Online stores servicing customers

● *Services of an agency/organisation/company where a number of clients are serviced at different locations:*

- Nurses visiting patients in their homes
- Salespeople visiting clients
- Repair-technicians servicing clients
- Social support services where volunteers visit vulnerable people.
- Advertising agents visiting prospective clients

- *Vehicle routing with real-time traffic information:*
  - Taxi car fleets servicing customers placing calls from different locations
  - Ambulance cars servicing calls from different locations
  - Traffic police cars dealing with accidents
- *Gas supply from different storage locations*
- *Electrical supply from different generators*
- *Water supply from different water reservoirs*
- *Data supply from interchangeable data sources*

  (e.g. Server streaming of video files from interchangeable websites to clients)

Consider now a number of interchangeable sources connected to a number of destinations. A *connection* is understood to be a unique sequence of edges, steps, lines, trajectories, etc. which provide a link between a source and a destination. Assume for the sake of simplicity that each source can service exactly one destination at a time and each destination can be serviced by exactly one source at a time. Typical examples are a fleet of several of taxi/ambulance/police cars servicing calls at different locations and a number of advertising agents visiting clients in their homes.

Figure 1a features logistic supply networks where a particular commodity or service is delivered from the three interchangeable sources *s*1,*s*2 1nd *s*3 to three destinations *t*1,*t*2, *t*3.

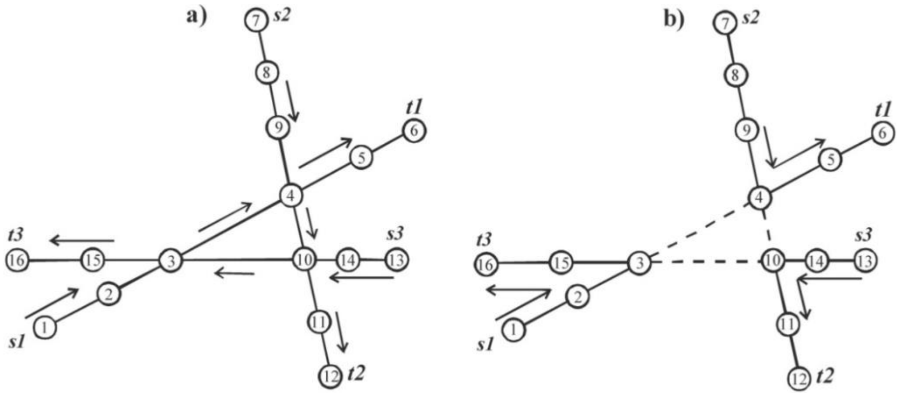

Figure 1. Augmenting a cyclic path negative cost in logistic supply networks.

Consider a network defined on undirected edges. The connecting paths between the sources s1,s2 and s3 and destinations d2,d2 and d3 is equivalent to transporting flow of unit magnitude between the sources s1,s2 s3 to the destinations *d1,d2 and d3* (Fig.1a,2a).

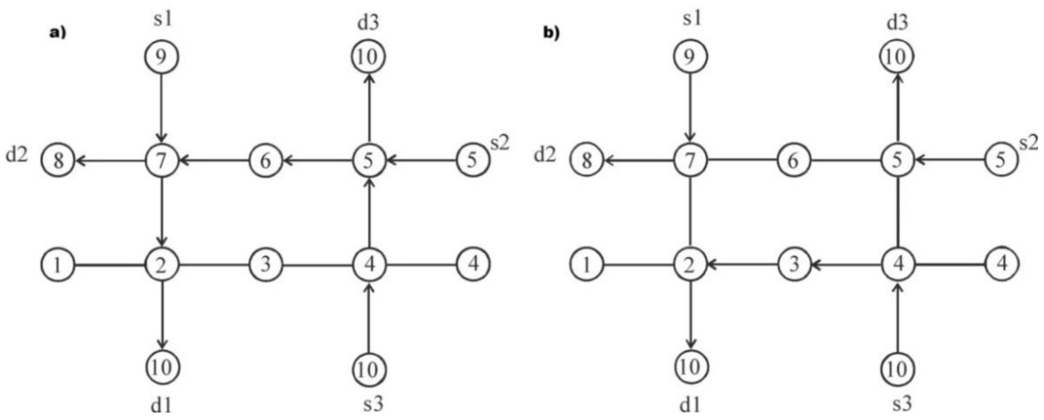

Figure 2. Augmenting the cyclic path (4,5,6,7,2,3,4) with negative cost reduces the total cost associated with the set of connections.

It is assumed that the capacity of the undirected edges is very large and cannot be exceeded by several connection/paths using the same edge simultaneously. The connection of a source and destination consists of *directed edges only* and will be referred to as *connecting path*.

When a new connection is made from a source s to a destination d, each edge with zero flow becomes an edge with forward flow equal to one unit. If an edge with backward flow is encountered the backward flow is decreased by one unit.

As a result, at each node k in the resultant flow network, the flow conservation is preserved: the sum of the flows $\sum_i f(i,k)$ entering node k is always equal to the sum of the flows $\sum_j f(k,j)$ exiting node k.

$$\sum_i f(i,k) = \sum_j f(k,j) \tag{6}$$

A *cyclic path* is a sequence of nodes i where the ith node is adjacent to the i+1st node and the last node coincides with the first node. A cyclic path is always associated with a specified direction of traversal. If the direction of the flow along an edge from the cyclic path coincides with the direction of traversal of the edge, the edge will be referred to as edge with forward flow or simply a *forward edge*. If the direction of the flow along an edge is in opposite direction to the direction of traversal of the edge, the edge will be referred to as edge with backward flow or simply *backward edge*. If the flow along an edge is zero, the edge will be referred to as *zero edge*.

Thus, the cyclic path 3,4,10,3 in Fig.1a contains forward edges only. The cyclic path 4,5,6,7,2,3,4 in Fig.2a contains forward and zero edges only, while the cyclic path 2,7,6,5,4,3,2 in Fig.2a contains backward and zero edges only. Zero edges and forward edges will be referred to as '*closing edges'*.

A cyclic path may not include any source or a destination, may include exactly one source or exactly one destination or may include both a source and a destination. The flow along a cyclic path can be increased by a certain value along all closing edges (forward edges and zero edges) and decreased by the same value along all backward edges. This process will be referred to as *flow augmentation*.

If a cyclic path includes a source, the edge coming out of the source can appear *only as a backward edge* during the flow augmentation because each source can serve exactly one destination at a time. If the cyclic path includes a destination, the edge going into the destination can appear only as a backward edge in the cyclic path because each destination can be serviced by exactly one source at a time. If these conditions are fulfilled, the cyclic path will be referred to as *augmentable cyclic path*.

5. COST FACTOR OF A CYCLIC PATH. PROPERTIES OF THE CONNECTIONS WITH THE INTERCHANGEABLE SOURCES. AUGMENTATION OF A CYCLIC PATH.

Suppose that a *detriment/harm* of a particular type is present along the sections (edges) building the connections from the sources to the destinations. The detriment can be a time delay due to congestions, cost of fuel spent on transportation, quantity of CO_2 emissions, usage of a vehicle, etc. If a particular edge is used by more than one connection, it is assumed that the detriment affects all paths using this edge and the total detriment along the edge is proportional to the number of connections using the edge.

Suppose that the detriment is defined as cost equal to c_i units along edge *i*. Augmenting a cyclic path with Δ units of flow means that the flow along each backward edge is decreased by Δ units and the flow along each closing edge is increased by Δ units.

Because, during the cyclic path augmentation, the flow along closing edges is increased, the cost along these edges will increase. For the *i*th closing edge, the increase of the cost ΔC_i associated with this edge is equal to

$$\Delta C_i = \Delta \times c_i^{(c)} \tag{7}$$

The increase of the cost related to the entire cyclic path composed of M_c closing edges only, is given by

$$\Delta C_c \approx \Delta \times \sum_{i=1}^{M_c} c_i^{(c)} \tag{8}$$

where $c_i^{(c)}$ is the cost along the *i*th closing edge.

Suppose that a path containing both backward and closing edges is augmented with flow of magnitude Δ. After the cyclic path augmentation, along any backward edge, the flow of magnitude Δ is no longer transported but is prevented from being transported, because the flow along backward edges has been *decreased* by Δ. Therefore, the increase in the cost ΔC_j, along a backward edge with index '*j*', has a negative sign:

$$\Delta C_j = -\Delta \times c_j^{(b)} \tag{9}$$

For a cyclic path including M_b backward edges, the total change ΔC_b of the cost can be approximated by

$$\Delta C_b \approx -\Delta \times \sum_{j=1}^{M_b} c_j^{(b)} \tag{10}$$

where $c_j^{(b)}$ is the cost associated with the *j*th backward edge.

Consequently, the total increase of the cost ΔC along a cyclic path including backward and closing edges can be approximated by

$$\Delta C = \Delta C_f + \Delta C_b \approx \Delta \times \left(\sum_{i=1}^{M_c} c_i^{(c)} - \sum_{j=1}^{M_b} c_j^{(b)} \right) \tag{11}$$

The quantity

$$\gamma = \sum_{i=1}^{M_c} c_i^{(c)} - \sum_{j=1}^{M_b} c_j^{(b)} \tag{12}$$

will be referred to as a *cost factor of a cyclic path*. As it will be shown later, this concept is of fundamental importance to reducing the risk associated with a set of connections from interchangeable sources to a set of destinations.

A positive sign of the cost factor γ for a particular direction of traversal means that the cost impact associated with this local modification is positive. In other words, if the cyclic path is augmented with flow in this direction of traversal, the total cost will increase. A negative sign of the cost factor γ for a particular direction of traversal means that the cost impact from this local modification is negative. The larger the magnitude of the cost factor γ, the larger is the amount of the cost increase/decrease following a cyclic path augmentation.

A *negative cyclic path* stands for a cyclic path characterized by a negative cost factor in one of the directions of traversal. The augmentation of a negative cyclic path results in a decrease of the total cost associated with the set of connections. At the same time, the connections from sources to destinations will not be affected. Effectively, augmenting negative cyclic paths is a process of minimizing local costs and reducing the total cost associated with the connecting paths. This is effectively a process of *drawing value from the re-optimization of the existing connections*.

For a set of interchangeable sources and a set of consumers, the following lemma holds.

Lemma 1. An augmentation of a cyclic path preserves the basic property: each source services exactly one destination and each destination is being serviced by exactly one source.

Proof: Because all n sources initially supply a unit flow to each of the n destinations, after an augmentation of a cyclic path, n sources will still originate n units of flow (one unit per source) and n destinations will still accept exactly n units of flow (one unit per destination).

In addition, at each node belonging to the cyclic path, the flow conservation will be preserved.

A process of path tracing can be started from one of the sources s. An edge coming out of the source can always be selected, carrying a unit flow towards a particular node i. According to the flow conservation law, there must be another edge incident to node i, carrying a unit flow which goes out of node i. By selecting this edge, the node i+1 is reached and so on. By continuing this process, the traced path visits either (i) an already visited node or (ii) a destination t (Fig.3).

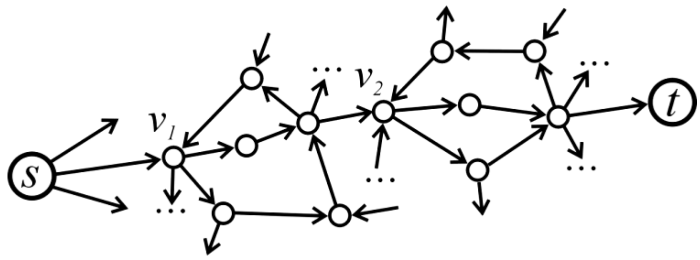

Figure 3 Tracing a connecting path from a source *s* to a destination *t*.

(i) Consider the first alternative. If a node v_1 has been visited again before the destination *t* has been reached (Fig.3), a directed flow cycle has been discovered in the network. A unit flow can then be removed from all edges of this cycle, which will result in a new feasible network flow where the flow conservation law holds at each internal node (different from a source or a destination). After removing the unit flow from all edges belonging to the directed flow cycle, the process continues from the repeated node $v1$. The continuation is guaranteed, because exactly one unit of flow entering node v_1 and exactly one unit of flow leaving node v_1 had been removed. Therefore, according to the flow conservation law, at node v_1 there must be a non-empty edge leaving the node. If another cycle is encountered at node v_2, a unit of flow is removed from all edges of the unit cycle and the process continues from node $v2$. By continuing this process, it is guaranteed that a destination *t* will finally be reached. This is guaranteed, *because the flow conservation law holds at each encountered internal node and the network has been saturated with a finite number of flow units*.

(ii) Consider now the second alternative where a destination *t* has been reached. The traversed edges, each carrying a unit flow, then constitute a directed s-t path and from each of the forward edges of this s-t path, exactly one unit of flow is removed. This procedure however, will subtract one unit of flow from the only non-empty outgoing edge from the source *s* and exactly one unit of flow from the only non-empty edge entering the destination *d*. As a result, the procedure will effectively disconnect the source *s* and the destination *d* from the network and will leave a network with *n*-1 sources and *n*-1 destinations.

Another source with non-empty outgoing edge can then be selected and the entire process of source-destination tracing is repeated until another source-destination match is found. This is guaranteed because the flow conservation law at the nodes has been preserved by the previous traversal. Continuing this procedure will match at the end each source to a destination with a directed connecting path.□

For a set of interchangeable sources and a set of consumers, the following theorem also holds:

Theorem 1. *The necessary and sufficient condition for a minimum total cost related to a set of connections between interchangeable sources and destinations is the non-existence of cyclic paths with negative cost*

This theorem follows directly from the principle of the same sign local modifications provided that the system is connected and with additive response. To ensure that the system is indeed connected and with additive response, a theorem, stated in Ahuja et al. (1993) will be used. Here, this theorem has been stated as Lemma 2.

Lemma 2. *In a flow network, if there are two different feasible edge flows* $f_1(i,j)$ *and* $f_2(i,j)$*, resulting in the same throughput flow, the flow* $f_2(i,j)$ *can always be presented as a sum of the flow* $f_1(i,j)$ *and the augmented flows along a set of augmentable cyclic paths.*

Lemma 2 essentially provides the assurance that the system is connected and with additive response. In other words, from an initial state S_1, the network can be brought to any other possible state S_i solely by a set of permissible local modifications consisting of augmenting cyclic paths.

State 1 corresponds to a network characterised by edge flows $f_1(i,j)$ and throughput flow equal to n (from n sources to n destinations). The total cost D_1, associated with edge flows $f_1(i,j)$ is assumed to be the smallest possible. If a state S_2 exists characterised by edge flows $f_2(i,j)$, associated the same throughput flow n and a smaller total cost $D_2 < D_1$, according to Lemma 2, state S_2 can be obtained from state S_1 after adding the augmented flows along k cyclic paths. The expected total cost D_2 associated with state S_2 is given by

$$D_2 = D_1 + \Delta_1 \times \gamma_1 + \Delta_2 \times \gamma_2 + ... + \Delta_k \times \gamma_k \quad (13)$$

where γ_i is the cost factor of the ith cyclic path and Δ_i is the flow with which the ith cyclic path is augmented.

The system is with additive response and the conditions of Theorem 1 are fulfilled. Consequently, if the system is brought to a state where there are no negative cost cyclic paths, the minimum global cost associated with the system of connections is guaranteed. The set Ω of permissible local modifications are the augmentations of the possible cyclic paths. A local modification results in positive local effect only if the augmentation of the cyclic path in any of the two possible directions of traversal results in a positive change of the cost. This is possible only if the cost factor of the cyclic path is positive in any of the directions of traversal. Permissible modifications with positive local effect only exist if no augmentable cyclic path associated with a negative cost factor exists.

Consequently, removing all cyclic paths with negative cost will result in a state associated with the smallest global cost associated with the set of connections. *Attaining the global minimum of the total cost has been guaranteed by a series of permissible local modifications*.

The next theorem guarantees that the process of augmentation of cyclic paths with negative cost, until no such cyclic path can be found, will end after a finite number of steps.

Theorem 2. *The process of augmenting cyclic paths with negative cost, until no such paths can be found, is finite and will end after a finite number of cyclic path augmentations.*

Proof.

The edges of the network connections can be thought as a flow network where each edge has been saturated with certain number of flow units. Because the costs along the edges can always be expressed as rational numbers, multiplying the costs along the edges by their common denominator will express all costs with integer numbers. Suppose that there are m edges and the maximum number of cost units characterising an edge is u_{max}. The total amount of cost units contained in the network therefore does not exceed mu_{max}.

Because only cyclic paths with negative cost are augmented, each cyclic path augmentation removes at least one cost unit from the total amount of mu_{max} cost units contained in the network. Consequently, after a finite number of steps, there will be a step at which no cyclic path with a negative cost factor will exist.

5.1 REDUCING RISK BY AUGMENTING CYCLIC PATHS WITH NEGATIVE COST FACTOR

Cyclic paths with negative cost factors are **highly undesirable** because they are associated with **wastage of energy and time and increased levels of congestion and pollution to the environment**. Cyclic paths with negative cost factor however, can be removed without affecting the throughput flow from sources to destinations. Optimising real flow networks with interchangeable sources is impossible without eliminating first all cyclic paths with negative cost factor.

Removing loops with a negative cost is the key to optimising any commodity supply network involving interchangeable sources. Optimising commodity supply is the key to developing the intelligent transportation and communication networks of the future.

In Fig.1a, by augmenting the cyclic path 3,4,10,3 with unit flow in opposite direction, the closed loop is eliminated (Fig.1b) without affecting the connections from the sources $s1$, $s2$ and $s3$ to the destinations $t1$, $t2$ and $t3$. The transportation losses, the risk of congestion and accidents have been reduced without affecting the connections from the sources to the destinations. As a result, value has been derived from significantly reducing the transportation losses, the risk of congestion, delay and accidents.

The connections do not have to form closed loops for a cyclic path with negative cost factor to be present. In Fig.2, the cyclic path (4,5,6,7,2,3,4) has a dominant flow (4,5,6,7,2) along the direction of traversal. The cyclic path can be augmented with flaw in opposite direction. As a result, the transportation losses are reduced without affecting the throughput flow from the sources $s1$,$s2$ and $s3$ to the destinations d1,d2 and d3 (Fig.2b). The cyclic path with a negative cost factor has been augmented and the transportation losses reduced, without affecting the throughput flow from the sources s1,s2 and s3 to the destinations d1,d2 and d3.

Surprisingly, *selecting the nearest available source to supply the destinations does not guarantee an optimal solution*.

In the network in Fig.4a with sources s1,s2,s3 servicing destinations d1,d2,d3, the nearest source to destination d1 is s1, the nearest source to destination d2 is s2 and the nearest remaining source to destination d3 is s3. In addition, all of the source-destination pairs have been connected with the shortest paths. Despite this the obtained solution is not the optimal solution. The cyclic path 11,5,6,3,8,12,11 is with a negative cost and by augmenting it with unit flow, the new set of connections in Fig.4b are obtained. Removing the cyclic path with a negative cost factor by augmenting it with unit flow results in the network in Fig.4b where no cyclic paths with negative cost factors are present. The throughput flow from the sources s1,s2 and s3 to the destinations d1,d2 and d3 remains the same while the transportation costs have been reduced significantly.

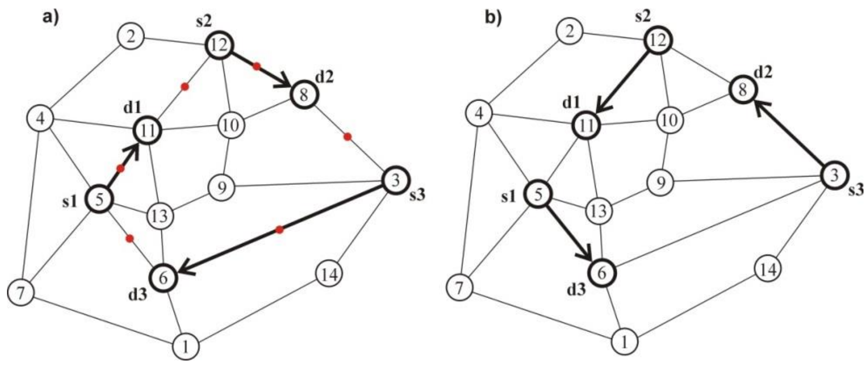

Figure 4. Selecting the nearest available source to supply the destinations along the shortest paths does not guarantee an optimal solution.

Cyclic paths with negative cost factors are also present in cases where the routes from sources to destinations are selected sequentially in time.

In the network from Fig.5a, the route (1,5,9,11,13) has been selected in the first hour and the route (2,8,9,12) has been selected in the next hour. Despite that the shortest source destination paths have been selected, cyclic paths with negative cost factors still appear. The cyclic paths with negative cost (1,3,12,9,5,1) and (2,10,13,11,9,8,2) in network (a) could be removed by augmenting them with unit flow, which yields network (b). The transportation costs have been reduced without affecting the throughput flow from the sources s1 and s2 to the destinations d1 and d2.

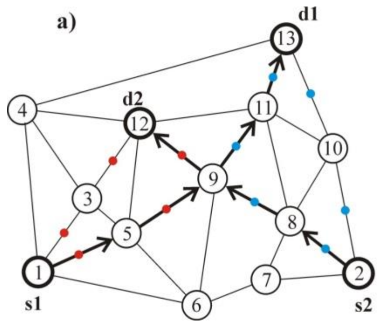

Figure 5. Negative cyclic paths are also present in cases where routes are selected sequentially in time

Cyclic paths with negative cost exist even if only a single connection is present.

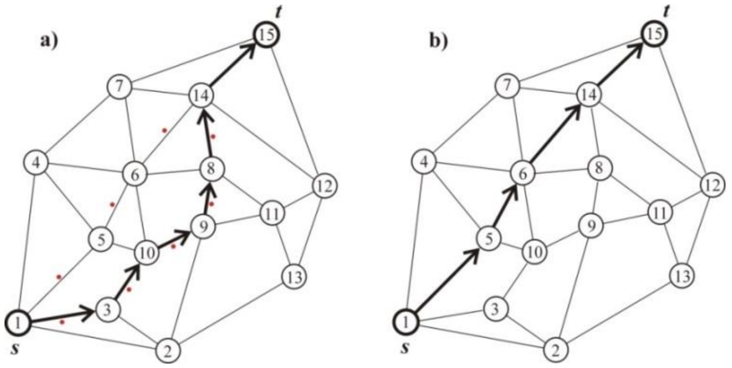

Figure 6 Negative-cost cyclic paths exists even if only a single flow path is present

Thus, in the network from Fig.6a, removing the loop with negative cost (1,5,6,14,8,9,10,3,1) results in network (b) where no loops with negative cost are present. The number of connections remains the same while the transportation costs have been reduced.

Cyclic paths with negative cost are associated with increased risk of congestion and accidents, wastage of energy, time, and increased levels of pollution to the environment. Removing such cyclic paths is the key to optimising supply networks involving interchangeable sources and to developing the intelligent transportation and communication networks of the future. Optimizing supply networks by draining highly undesirable loops with negative cost derives significant value, reduces the transportation costs, the risk of congestion, the number of accidents, and the environmental pollution. The result is billions of dollars saved to the world economy.

The existence of loops with negative cost remained unnoticed by scientists for nearly 60 years. Ironically, despite the years of intensive research on static flow networks, loops with negative cost appear even in the "network flow solutions" of all published algorithms (including the famous Ford-Fulkerson algorithm; Ford and Fulkerson, 1956) for maximizing the throughput flow in networks, since the creation of the theory of flow networks in 1956.

CONCLUSIONS

1. A simple generic risk-reduction principle with unexpected powerful applications has been formulated for systems with additive response with respect to a defined set of local modifications through which any possible system state can be reached. A sufficient condition for a global minimum for such systems is the positive sign of the local effects bringing the system from a particular initial state to any possible system state.

Similarly, the sufficient condition for a global maximum is the negative sign of the local effects bringing the system from a particular initial state to any possible system state.

3. The formulated principle, similar to the pigeonhole principle in combinatorics, has deep and non-trivial applications. It often provides an efficient strategy for attaining a global extremum for connected systems with additive response by conducting a series of permissible local modifications, which is often a simple task, achievable within a finite number of steps.

4. A real-life application of the formulated principle has been found in determining the global extremum of multivariable functions whose partial derivatives maintain the same sign in a rectangular domain.

5. The proposed principle can be applied with success to improve the performance of a heuristic algorithms searching for a global extremum which incorporate a local search to determine the local best value. The time complexity of the critical step is

reduced from exponential $O(2^n)$ to linear $O(n)$ which improves significantly the performance of the global optimisation algorithm.

6. The formulated simple generic principle has also been applied to common real-life systems including a set of interchangeable sources connected to a set of destinations. It is shown that the global transportation cost attains the global minimum when the flow augmentation along any cyclic path results in an increase of the transportation cost.

7. On the basis of a counter-example it has been shown that selecting the nearest available source to supply destinations along the shortest available paths, does not guarantee an optimal solution (the smallest global transportation cost). This is a rather unexpected result, contrary to well-established practices in network optimisation. Removing the parasitic flow loops in the network however, does guarantee an optimal solution.

8. The same sign local effects principle holds even if the magnitudes of the effects from the local modifications depend on the current levels of the input variables. This has been demonstrated with a function of several variables.

9. On the basis of a counter-example, it is demonstrated that the successive shortest-path algorithm does not guarantee a minimal transportation cost. Removing cyclic paths with negative cost is the key to optimising supply networks involving interchangeable sources and to developing the intelligent transportation and communication networks of the future.

10. Eliminating loops with a negative cost factor in logistic supply networks leads to a significant reduction of the risk of congestion and delays. The expected reduction of transportation costs and the environmental pollution has the potential to save billions of dollars to the world economy.

REFERENCES

Altshuller, G.S. (1999). *The Innovation Algorithm, TRIZ, Systematic Innovation and Technical Creativity*, Technical innovation Center, Inc. Worcester.

Altshuller, G.S. (1996). *And suddenly the inventor appeared, TRIZ, the theory of Inventive Problem Solving*, Translation from Russian, 1996.

Altshuller, G.S. (1984). Creativity as an exact science: The theory of the solution of inventive problems, Gordon and Breach Science Publishing, New York.

Anderson, T.L. (2005). Fracture Mechanics: Fundamentals and Applications, Taylor & Francis.

Aven T. (2003). *Foundations of risk analysis*, Wiley.

Barlow, R.E. and Proschan F. (1975). *Statistical theory of reliability and life testing*, Rinehart and Winston, Inc..

Bedford, T. and Cooke R. (2001). *Probabilistic risk analysis, foundations and methods*, Cambridge University Press.

Billinton, R. and Allan R.N. (1992). *Reliability evaluation of engineering systems*, 2nd ed., Plenum press.

Bird, G.C. and D. Saynor. (1984). 'The effect of peening shot size on the performance of carbon-steel springs', *Journal of Mechanical Working Technology* vol.10 (2) 175-185.

Brualdi, R.A. (2010), *Introductory Combinatorics* (5th ed.), Pentice Hall.

Collins, J.A. (2003). *Mechanical design of machine elements and machines*, John Wiley & Sons.

Dhillon, B.S., C.Singh (1981). *Engineering reliability: New techniques and applications*, John Wiley & Sons.

Ebeling, C.E. (1997). *An introduction to Reliability and Maintainability Engineering*, McGraw-Hill.

Edmonds J. and R.M.Karp, Theoretical improvements in algorithmic efficiency for network flow problems, Journal of the ACM, 19(2) (1972) pp.248-264.

Ewalds, H.L. and Wanhill, R.J.H. (1984). *Fracture Mechanics*, Edward Arnold, London.

Ford, L.R. and D.R. Fulkerson, Maximal flow through a network (1956). *Canadian Journal of Mathematics*, vol.8(5), pp.399-404.

French, M. (1999). *Conceptual design for engineers*, 3rd ed., Springer-Verlag London Ltd.

Hertzberg, R.W. (1996). *Deformation and fracture mechanics of engineering materials*, 4th ed., John Wiley & Sons, Inc.

Kiranyaz S., T.Ince, M.Gabbouj, Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition, Springer, 2014.

Lazinica A (ed.). *Particle Swarm Optimization*, In-Tech, 2009.

Lewis, E.E. (1996). *Introduction to reliability engineering*, John Wiley & Sons, Inc.

McKinsey Global Institute, Global flows in a digital age: How trade, finance, people, and data connect the world economy, report, McKinsey & Company 2014, April 2014.

Pahl G., W.Beitz, J.Feldhusen, K.H.Grote (2007). *Engineering design*, Springer.

Phadke, M.S. (1989). *Quality engineering using robust design*, Prentice Hall, Englewood Cliffs.

Niku-Lari, A. (1981). 'Shot-peening', *In* the *First International Conference on Shot Peening*, Paris, 14–17 September, 1-27, Pergamon Press.

O'Connor, P.D.T (2003). *Practical reliability engineering*, 4th ed., John Wiley & Sons.

Ramakumar, R. (1993). *Engineering reliability, fundamentals and applications*, Prentice Hall.

Sundararajan, C (Raj) (1991). *Guide to reliability engineering: Data analysis, applications, implementations and management*, Van Nostrand Reinold.

Tanenbaum A.S., *Computer networks*, 4th ed., Pearson Education International, (2003).

Vose D., Risk Analysis (2002), A quantitative guide, John Wiley & Sons.

Zahavi E., Torbilo V. (1996). *Fatigue design, Life expectancy of machine parts*, CRC Press, 1996.