Economic Interpretation of Demand Curves in Multi-product Electricity Markets

Feng Zhao, Tongxin Zheng, Eugene Litvinov

February 5, 2020[1]

Abstract

In the absence of demand-side bids for certain reliability products in the wholesale electricity markets, Independent System Operators (ISOs) traditionally use fixed demand requirements with penalty factors to clear the market. This approach does not allow proper tradeoffs between cost and reliability due to the inelasticity of the fixed requirements. Therefore, ISOs have been replacing the fixed requirements with sloped demand curves. However, the economic interpretation of demand curves, especially for multiple coupling products, are largely missing in both theory and practice. This could lead to the misrepresentation of the demand benefit and the market clearing objective. Using a two-product market model, this paper reveals two distinct interpretations of demand curves, each associated with a specific form of the market clearing problem. This implies that the construction of demand curves and the use of them in market clearing must be consistent with their economic interpretation. We also analyze the sequential and iterative market clearing processes currently implemented in some markets to demonstrate their solution sub-optimality. The root cause of the problem is the lack of a clear representation of the demand benefit. The social surplus optimization based on the proper interpretation of demand curves is proposed.

Key words: Demand curve, economic interpretation, electricity markets, fixed requirement, iterative clearing, penalty factor, reliability, sequential clearing, social surplus optimization.

Feng Zhao (fzhao@iso-ne.com), Tongxin Zheng (tzheng@iso-ne.com), Eugene Litvinov (elitvinov@iso-ne.com) are with Business Architecture & Technology department of ISO New England, Inc., Holyoke, MA, USA.

The views expressed in this paper are those of the authors and do not represent the view of ISO New England.

[1] Original version uploaded on February 5, 2020, and last modified on February 26, 2020.

I. INTRODUCTION

In economics, demand curves play a critical role in forming market equilibria and clearing prices. Individual demand curves for power system reliability products such as capacity and reserve, however, are often absent in the wholesale electricity markets managed by Independent System Operators (ISOs) due to the "public good" nature of these products. As a result, ISOs often set fixed requirements for these products and purchase them from qualified suppliers (e.g., generation resources) in the wholesale markets. Moreover, when there is a shortage of supply to meet the requirements, ISOs often use the prescribed penalty factors to set market clearing prices. Such practice is equivalent to having a single-block demand curve for the corresponding product as shown in Fig. 1.

A simplified market clearing problem with a fixed requirement and a penalty factor can be formulated as:

$$\min_{Q,\delta} C(Q) + PF \times \delta \quad (1)$$

$$s.t. \quad Q \geq Req - \delta. \quad (2)$$

$$Q \in \Omega,\ \delta \geq 0. \quad (3)$$

where

$Q, C(Q)$: Cleared supply quantity and cost, respectively;

δ, PF: Slack variable and the penalty factor, respectively;

Req: Fixed requirement for the product;

Ω: Feasible region of the supply variable Q.

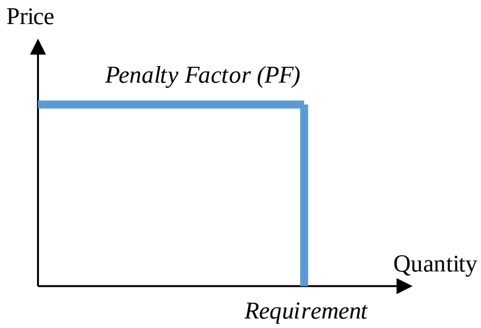

Fig. 1. Illustration of the fixed requirement and penalty factor.

The above model with the fixed requirement and the penalty factor suffers several drawbacks. *First*, the fixed product requirement does not allow the tradeoff between cost and reliability. As a result, the system could be unreliable if the ISO underestimates the requirement, or the ISO may overbuy the reliability at an excessive cost. *Second*, the clearing price tends to be volatile with the high penalty price for the supply shortage and zero price for the excessive supply, posing risks for the generation investment. *Lastly*, the approach does not recognize the reliability value of a product beyond the fixed requirement.

In view of the above drawbacks, most ISOs have been replacing the single-block penalty factors with multiple blocks, or more generally, sloped demand curves. For example, ERCOT has been using hour-ahead reserve forecast error distributions to derive real-time reserve demand curves [1]-[2]; NYISO has implemented multi-segment capacity demand curves in its spot capacity market [3]; PJM has been using capacity demand curves in its Base Residual Auction (BRA) [4] and recently proposed reserve demand curves for its day-ahead and real-time markets [5]; and ISO-NE implemented its system and local capacity demand curves based on the marginal reliability impact of capacity [6]. A typical sloped demand curve is illustrated in Fig. 2.

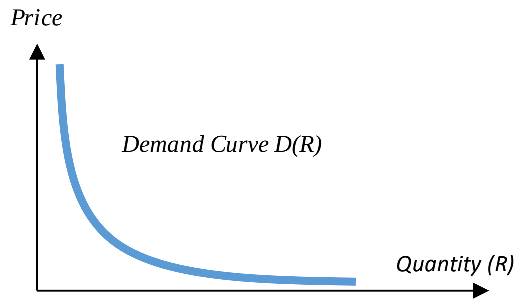

Fig. 2. An illustrative demand curve.

For comparison with the fixed requirement approach, the mathematical model of a single-product market clearing problem with the demand curve is presented as

$$\min_{Q,R} C(Q) - \int_0^R D(R)\, dR \tag{4}$$

$$s.t. \quad Q \geq R \tag{5}$$

$$Q \in \Omega,\ R \geq 0. \tag{6}$$

where $R, D(R)$ are the demand quantity and the demand curve, respectively.

Most of the demand curve related efforts, however, have focused on the shape or certain parameters of the curves without addressing a more fundamental question: *What is the economic interpretation of demand curves* in the market without consumers explicitly expressing their willingness to pay for the corresponding product? Such a question is especially important in a multi-product market where the demand curves for these products are inherently related. For example, a capacity market with local capacity zones requires system and zonal capacity demand curves, and a reserve market with 10-min and 30-min reserve products requires both 10-min and 30-min demand curves. The meaning of these interacting demand curves has to be clearly defined so that they can be appropriately constructed and used in market clearing.

Our previous work [7] on constructing capacity demand curves for ISO-NE leads to the definition of a local import zone's capacity demand curve as the "additional" marginal reliability benefit of the zonal capacity over the capacity outside the zone. The paper focused on how to construct capacity demand curves rather than the general interpretation of demand curves. To the best of our knowledge, no existing literature has systematically discussed **the meaning of interacting demand curves in a multi-product electricity market**. In this paper, we use a stylized two-product market model to analyze the meaning and the relation of the interrelated demand curves. It is revealed that different formulations of the market clearing problem may yield very different interpretations of the demand curves. Furthermore, we show that a clear interpretation of demand curves is essential for forming the demand benefit and the market clearing objective. We demonstrate this by examining a sequential clearing procedure and an iterative clearing algorithm currently implemented in some markets. Our analysis shows that these clearing methods lack a clear interpretation of demand curves and therefore may produce socially suboptimal solutions. The social surplus optimization formed on the proper interpretation of demand curves is proposed for market clearing.

The main contributions of this paper are: 1.) Identified the importance of demand curve interpretations in a multi-product market; 2.) Revealed different interpretations of demand curves under different formulations of the market clearing problem; 3.) Demonstrated the flaws in the existing sequential and iterative clearings of multi-product markets, and proposed the appropriate market clearing formulation for optimal social surplus.

The rest of the paper is organized as follows: *Section II* presents a stylized market clearing problem with two coupling products, and analyzes two different formulations and their distinct interpretations for demand curves. *Section III* and Section *IV*, respectively, present a sequential clearing algorithm and an iterative algorithm that are adopted in some existing markets. The potential sub-optimality issue of these algorithms is analyzed and attributed to the lack of clear interpretation of demand curves. The social surplus optimization problem is formulated. *Section V* presents numeric examples to compare solutions of the sequential/iterative algorithms and the proposed social surplus optimization. *Section VI* concludes the paper.

II. Economic Interpretation of Demand Curves

With the fixed requirement, any violation of the requirement is penalized by the penalty factor as shown in (1)-(3). Each penalty factor simply represents the market administrator's perceived price of violating the corresponding constraint. Penalty factors serve two functions:

- "***Feasibility***" function, i.e., to always ensure a solution from the market clearing software. While it is possible that physical resources cannot meet the requirement (i.e., physically infeasible), with the slack variable the problem of (1)-(3) will always yield a solution. The value of the penalty factor is often set high enough (e.g., greater than the offer price of any physical resource), so that a zero slack variable indicates the physical feasibility, and vice versa.
- "***Pricing***" function, i.e., to produce market clearing prices when the requirement cannot be met by physical resources. In this situation, penalty factors will set the market clearing prices. To avoid price spikes, in practice some markets would use a separate set of smaller penalty factors for pricing.

The administratively set values of penalty factors could be controversial, in particular when serving the "*pricing*" function, e.g., suppliers would favor a higher penalty factor while consumers would prefer a lower penalty factor. Furthermore, the feasibility function requires high-enough penalty factors, which conflicts the practical pricing needs for lower penalty factors. Nevertheless, the "*pricing*" function of penalty factors is only performed under the shortage of physical resources. Most of the time, penalty factors fulfill the less controversial "*feasibility*" function.

When demand curves replace penalty factors, treating the transition as a simple upgrade of penalty *values* to more sophisticated penalty *functions* is tempting. Therefore, stakeholder discussions often fail to address the meaning of demand curves and its implication on the market clearing. Instead, demand curve discussions have constantly centered on the curve's shape or parameters. The implicit assumption for this is that the existing market clearing process would not need to change with demand curves replacing penalty factors.

Indeed, the implementation of demand curves makes the "*feasibility*" function obsolete since there is no longer a fixed requirement to violate. In contrast, "*pricing*" becomes the primary function of demand curves. This functional shift makes the economic interpretation of demand curves crucial for the formation of the market clearing problem and its resulting clearing prices.

In the conventional economic theory, a demand curve, as illustrated in Fig. 2, represents the marginal benefit of purchasing a product. The demand benefit as the integral of the demand curve, together with the supply cost, form the market clearing objective of maximizing the social surplus, i.e., (*demand benefit – supply cost*). Such interpretation is intuitive in a single-product setting, but it is not obvious for multiple products with the corresponding demand curves interacting with each other. Questions to be answered include: Does a demand curve associated with each product represent the marginal benefit of purchasing that product? How would the interpretation reflect the couplings among multiple products? Is the interpretation relevant to the market clearing process?

In the following, we reveal that the interpretation of demand curves in a multi-product market depends on the market clearing formulation. Two formulations for a stylized two-product market clearing problem are presented in subsections II.1 and II.2, respectively, each implying a distinct meaning of demand curves. The results suggest that a proper interpretation of demand curves should dictate how the demand curve is constructed and how the market is cleared. In subsection II.3, we compare the two demand curve interpretations and then follow with the discussion on practical implications in subsection II.4.

Without loss of generality, in this paper we consider a market with two coupled products, i.e., a high-quality product *A* and a low-quality product *B*. The market has supply offers but no demand bids from participants. Such situation typically occurs in an ISO's one-sided markets such as capacity and ancillary service markets due to the lack of direct demand bids for these products. The coupling relation is characterized by the high-quality product *A*'s substitutability for the low-quality product *B*. In practice, these coupled products could be local and system capacity products, or hierarchical ancillary service products.

II.1 Demand curve as the marginal benefit of each individual product

In the single-product setting, the demand curve can be interpreted as the demand's marginal benefit of purchasing the product. A natural extension of this interpretation to the two-product setting is: **the demand curve for product *A* represents the marginal benefit of *A*, while the demand curve for product *B* represents the marginal benefit of *B*.** This interpretation can be conceived as if the consumers were to submit demand bids for the two products, with the demand bid curve for each product representing the marginal willingness to pay for that product. Then the total benefit of purchasing the two products is the sum of the individual product benefits as illustrated in Fig. 3.

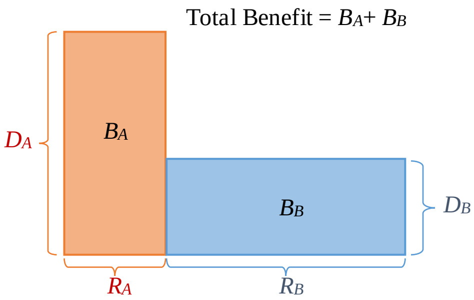

Fig. 3. Total benefit as the sum of individual product benefits.

In the above figure, R_A and R_B, respectively, are the demands for the two products. D_A and D_B, respectively, are the marginal benefits of the two products (showed as constant in the figure for simplicity of the illustration). The individual product benefits B_A and B_B, illustrated as colored areas in the above figure, are the integrals of the marginal benefit functions D_A and D_B, respectively. The total benefit of the two products as the sum of the individual product benefits thus appears in the social surplus optimization objective of the following market clearing problem:

$$\min_{\substack{Q_A,Q_B,\\R_A,R_B}} C_A(Q_A) + C_B(Q_B) - \int_0^{R_A} D_A(R)\,dR - \int_0^{R_B} D_B(R)\,dR \qquad (7)$$

$$s.t. \qquad Q_A \geq R_A \qquad (8)$$

$$Q_A + Q_B \geq R_A + R_B \qquad (9)$$

$$Q_A \in \Omega_A,\ Q_B \in \Omega_B \qquad (10)$$

$$R_A, R_B \geq 0. \qquad (11)$$

where

Q_A, Q_B: Cleared supply quantity of *A* and *B*, respectively;

$C_A(\cdot), C_B(\cdot)$: Supply costs of products *A* and *B*, respectively;

Ω_A, Ω_B: Supply feasibility regions of *A* and *B*, respectively;

R_A, R_B: Cleared demand of products *A* and *B*, respectively;

$D_A(\cdot), D_B(\cdot)$: Demand curves of *A* and *B*, respectively.

In the above, (8)-(9) represent the requirements for products *A* and *B*, respectively. As shown in (9), the extra supply $(Q_A - R_A)$ for the high-quality product *A*, together with the supply Q_B for the low-quality product *B*, are used to meet the demand R_B for product *B*. Next, we define the market clearing prices and show that these prices are consistent with the above interpretation of demand curves.

Consider that the above clearing problem is convex and the shadow prices of (8) and (9) are λ_A and λ_B ($\lambda_A, \lambda_B \geq 0$), respectively. Then the Market Clearing Prices (MCPs) for products *A* and *B* are defined as

$$MCP_A \equiv \lambda_A + \lambda_B. \qquad (12)$$

$$MCP_B \equiv \lambda_B. \qquad (13)$$

Note that the price cascading relation $MCP_A \geq MCP_B$ holds as a result of product *A* being substitutable for product *B* in meeting *B*'s requirement in (9).

Based on the KKT optimality condition of (7)-(11), we have

$$D_A\left(R_A^{opt}\right) = \lambda_A + \lambda_B, \tag{14}$$

$$D_B\left(R_B^{opt}\right) = \lambda_B. \tag{15}$$

where R_A^{opt} and R_B^{opt} are the optimally cleared demands of products *A* and *B*, respectively. From (12)-(15), we have

$$D_A\left(R_A^{opt}\right) = MCP_A, \tag{16}$$

$$D_B\left(R_B^{opt}\right) = MCP_B. \tag{17}$$

Since the market clearing price for a product reflects the marginal benefit of demand for the product, (16)-(17) indicate that the demand curve for each product represents the marginal benefit of demand for that product, which is consistent with the interpretation of demand curves in this subsection.

II.2 Demand curve as additional marginal benefit of one product over another

In the previous subsection, the total demand benefit is represented as the sum of individual product benefits (see Fig. 3). Since the high-quality product *A*'s marginal benefit is greater than the low-quality product *B*'s, the total benefit can also be represented as the basic benefit of both products providing low-quality service (i.e., area B_C), plus the additional or premium benefit of the high-quality product *A* (i.e., area B_A') as illustrated in Fig. 4.

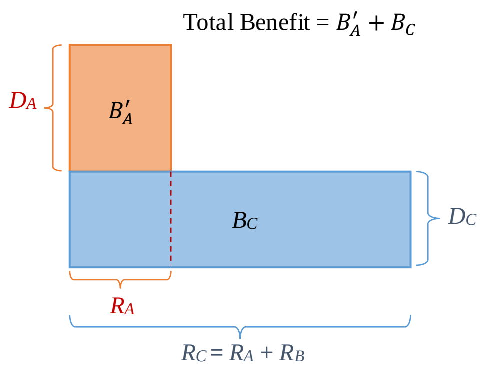

Fig. 4. Total benefit as the sum of basic and additional benefit.

Under the benefit composition in Fig.4, the total demand for both products is denoted by R_C ($=R_A + R_B$). **Demand curve D_C represents the "basic" marginal benefit of both products,** and **demand curve D_A' represents the "additional" marginal benefit of the high-quality product *A* over the basic marginal benefit.** The overall benefit of purchasing the two products is the sum of the integrals of the two demand curves (i.e., $B_A' + B_C$). The corresponding market clearing problem is then represented by:

$$\min_{\substack{Q_A,Q_B,\\R_A,R_C}} C_A(Q_A) + C_B(Q_B) - \int_0^{R_A} D_A'(R)\,dR - \int_0^{R_C} D_C(R)\,dR \tag{18}$$

$$s.t. \quad Q_A \geq R_A \tag{19}$$

$$Q_A + Q_B \geq R_C \tag{20}$$

$$Q_A \in \Omega_A,\ Q_B \in \Omega_B \tag{21}$$

$$R_A, R_C \geq 0. \tag{22}$$

where

R_C: Total demand for products *A* and *B*;
D_A', D_C: Demand curves associated with R_A and R_C, respectively.

Compared to (7)-(11), the main difference in the above formulation (18)-(22) lies in the model of variables and constraints related to product *B*, i.e., R_B in (9) reflects the demand of a single product *B*, while R_C in (20) reflects the overall demand for both products. Next, we define the market clearing prices and show that these prices are consistent with the above interpretation of the demand curves.

Suppose that the above problem is convex and the shadow prices associated with the product requirement constraints (19)-(20) are λ'_A and λ_C, respectively. Then the market clearing prices for products *A* and *B* are defined as follows:

$$MCP_A \equiv \lambda'_A + \lambda_C, \quad (23)$$

$$MCP_B \equiv \lambda_C. \quad (24)$$

Note that the price cascading relation $MCP_A \geq MCP_B$ holds. Based on the KKT optimality conditions of (18)-(22), we have

$$D'_A\left(R_A^{opt}\right) = \lambda'_A\ , \quad (25)$$

$$D_C\left(R_C^{opt}\right) = \lambda_C. \quad (26)$$

where R_A^{opt} and R_C^{opt} are the optimally cleared demand of *A* and the total demand, respectively. From (23)-(26), we have

$$D'_A\left(R_A^{opt}\right) = MCP_A - MCP_B, \quad (27)$$

$$D_C\left(R_C^{opt}\right) = MCP_B. \quad (28)$$

Since the market clearing price of a product represents the marginal benefit of consuming the product, or the demand's willingness to pay for the product, therefore (27) indicates that demand curve $D'_A(\cdot)$ reflects the "*additional"* willingness to pay for product *A* over the basic product, and (28) indicates that demand curve $D_C(\cdot)$ represents the willingness to pay for the basic product *B*–level service provided by both products. Such indication is consistent with the interpretation of demand curves in this section.

II.3 Comparison of Demand Curve Interpretations

The aforementioned different interpretations of demand curves arise from different compositions of the total demand benefit illustrated in Fig. 3 and Fig. 4. A basic question arises: **Are these two benefit compositions in subsections II.1 and II.2 equivalent**? In other words, can one demand benefit representation be translated into another equivalently? The following theorem answers the above question:

Theorem 1. The total demand benefit as the sum of individual benefits of the two products, is equivalent to the sum of the basic benefit of both products and the additional benefit of the high-quality product *A*, *if and only if* the basic benefit demand curve function D_C is constant. Namely,

$$\int_0^{R_A} D_A(R)\,dR + \int_0^{R_B} D_B(R)\,dR = \int_0^{R_A} D'_A(R)\,dR + \int_0^{R_C} D_C(R)\,dR,\ \ \forall (R_A, R_B)\ \ iff\ \ D_C(\cdot)\ \text{ is constant.}$$

[*Proof*]. Substituting out the cleared system demand R_C by the sum of cleared demands for both products (i.e., $R_A + R_B$) in the above equation yields

$$\int_0^{R_A} D_A(R)\,dR + \int_0^{R_B} D_B(R)\,dR = \int_0^{R_A} D'_A(R)\,dR + \int_0^{R_A+R_B} D_C(R)\,dR,\ \ \forall (R_A, R_B) \quad (29)$$

Taking derivatives over R_A and R_B, respectively, to both sides of the above equation yields

$$D_A(R_A) = D'_A(R_A) + D_C(R_A + R_B)\ ,\ \forall (R_A, R_B) \quad (30)$$

$$D_B(R_B) = D_C(R_A + R_B),\ \forall (R_A, R_B) \quad (31)$$

The above two equations hold for any values of R_A and R_B if and only if $D_C(\cdot)$ is a function of neither R_A nor R_B, i.e., $D_C(\cdot)$ is constant. [*End of Proof*]

Note that if $D_C(\cdot)$ is constant, then $D_B(\cdot)$ must be the same constant according to (31). The above theorem indicates that unless the basic marginal benefit function $D_C(\cdot)$ or the demand curve $D_B(\cdot)$ for the lower quality product *B* is constant, the two benefit compositions under the corresponding demand curve interpretations in subsection II.1 and subsection II.2 are not interchangeable in general. Therefore, a natural follow-up question is: **Which demand curve interpretation is better?**

Under both demand curve interpretations, the total benefit is considered to be additively separable in terms of the two demand variables, e.g., R_A and R_B in subsection II.1, or R_A and R_C in subsection II.2. Such consideration might be a necessary simplification for the demand curve design and the market clearing. However, in reality the actual total benefit function may not be separable. Therefore, any benefit composition such as the two discussed in subsections II.1 and II.2 can be viewed as an approximation to the actual non-separable benefit function. Then the answer to the above question would depend on whether or not the benefit composition under a demand curve interpretation is a good approximation of the actual benefit function. The answer would also depend on the complexity of the demand curve construction and the market clearing under a particular demand curve interpretation.

In sum, **the choice of demand curve interpretation is system and design dependent and there is no universal answer**. For instance, in ISO-NE's Forward Capacity Market (FCM), the avoided Loss of Load (LOL) is used to measure the reliability benefit of capacity. The demand curve interpretation described in subsection II.2 was adopted since the corresponding reliability benefits formed as (18) is found to be a good approximation of the actual LOL, and the concept of the total system capacity need R_C (instead of the rest-of-system need R_B) and its reliability evaluation process had already existed prior to the implementation of demand curves.

II.4 Practical Implications of Demand Curve Interpretations

It can be seen from the above analysis that demand curves could have different economic interpretations, each tied to a specific market clearing model. As a result, a clear and proper interpretation of demand curves is crucial since it dictates how the curves should be constructed and used in the market clearing process. However, demand curves are often viewed as merely an upgrade from penalty factors, and discussions have focused on the curve parameters instead of the fundamental economic interpretation of the curves. This brings a potential problem that the curve's construction could be inconsistent with the market clearing model. For instance, a market with local (high-quality) and system (low-quality) products may inherit from the penalty-factor era the clearing model of (18)-(22) that interprets the local demand curve as the "*additional*" marginal demand benefit. However, if the local demand curve is constructed to reflect the cost of a local marginal unit (instead of the cost additional to that of a marginal unit in the rest-of-system), then the clearing process based on (18)-(22) will overvalue the local demand's benefit and thus may over clear the zonal product or under clear the rest-of-system product. In sum, it is critical to have **the construction and the use of demand curves be consistent with their interpretation**.

Furthermore, in practice, the demand curves constructed without recognizing their economic meaning may not fit exactly into either of the two interpretations described in subsections II.1 and II.2. For instance, under the two-product setting, a market may have one demand curve for the high-quality product A that can be interpreted as D_A in subsection II.1, and another demand curve for both products that can be interpreted as D_C in subsection II.2. Such situation appears in some existing markets. Then the simple sum of the integrals of these two demand curves, would double count the basic benefit of product A as illustrated by the overlapped area B_{AC} in Fig. 5.

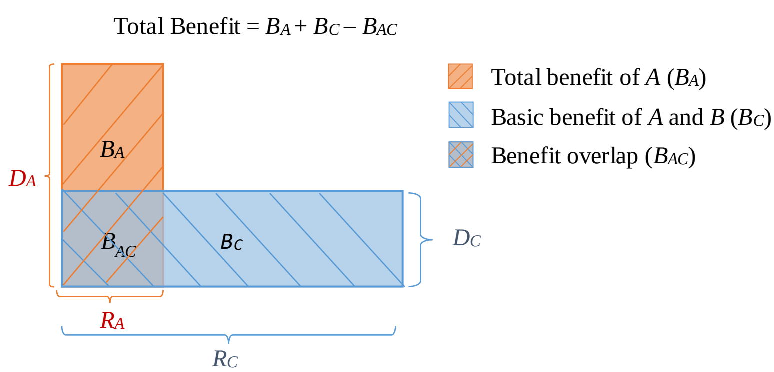

Fig. 5. Total benefit as the sum of B_A and B_C minus the overlap B_{AC}.

In the above figure, the total benefit (i.e., basic and premium) associated with the high-quality product A's demand R_A, i.e., B_A, can be mathematically represented by the integral of the product's demand curve D_A, while the basic benefit associated with the total demand R_C of both products, i.e., B_C, can be represented by the integral of the demand curve D_C. Note that the demand curves shown in the figure are constant only for the simplicity of illustration.

To form the total benefit, the overlapped benefit B_{AC} needs to be subtracted from the sum of integrals, i.e.,

$$Total\ Benefit = B_A + B_C - B_{AC}$$
$$= \int_0^{R_A} D_A(R)\,dR + \int_0^{R_C} D_C(R)\,dR - B_{AC}. \qquad (32)$$

To evaluate (32), additional assumptions will be needed on the form of B_{AC}. Essentially, B_{AC} represents the basic benefit associated with the demand R_A of the high-quality product *A*. Denote B_B as the basic benefit associated with the demand R_B of the low-quality product *B*. Then the basic benefit B_C associated with the total demand R_C $(= R_A + R_B)$ of both products is represented by

$$B_C = \int_0^{R_C} D_C(R)\,dR = \int_0^{R_A+R_B} D_C(R)\,dR = B_{AC} + B_B. \qquad (33)$$

So the form of B_{AC} depends on how the total basic benefit of $(R_A + R_B)$ is distributed between product *A* with quantity R_A and product *B* with quantity R_B. Below we present two assumptions of B_{AC}, which result in the demand curve interpretations in subsections II.1 and II.2, respectively.

Assume that the basic benefit of product *A* is associated with the high-value part of the monotonically decreasing demand curve D_C, i.e.,

$$B_{AC} = \int_0^{R_A} D_C(R)\,dR. \qquad (34)$$

Substituting (34) into (32), we have

$$Total\ Benefit = \int_0^{R_A} \big(D_A(R) - D_C(R)\big)\,dR + \int_0^{R_C} D_C(R)\,dR. \qquad (35)$$

Define $D_A'(R) \equiv D_A(R) - D_C(R)$. Then D_A' can be viewed as product *A*'s additional marginal benefit over its basic benefit. Then D_A' and D_C have the demand curve interpretation of subsection II.2.

Alternatively, *assume* that the basic benefit of product *A* is associated with the low-value part of the monotonically decreasing demand curve D_C, i.e.,

$$B_{AC} = \int_{R_C-R_A}^{R_C} D_C(R)\,dR. \qquad (36)$$

Substituting (36) into (32) and with $R_B = R_C - R_A$, we have

$$Total\ Benefit = \int_0^{R_A} D_A(R)\,dR + \int_0^{R_B} D_C(R)\,dR. \qquad (37)$$

Define $D_B(R) \equiv D_C(R)$. Then D_A and D_B have the demand curve interpretation of subsection II.1.

It can be seen from the above that the two assumptions of (34) and (36) on B_{AC} lead to different interpretations of demand curves. Assumption (34) assigns more demand benefits to product *A*, while (36) allocates more demand benefits to product *B*. Therefore, given the demand curve of product *A* and the supply offers, the assumption (34) tends to have less product *B* cleared as compared to the assumption (36). This demonstrates the importance of having a clear interpretation of demand curves and demand benefit function. Nevertheless, in some markets, a sequential or an iterative process without a clear interpretation of demand curves is adopted as the market clearing algorithm, which obscures the form of the total benefit function (32) and the subsequent social surplus optimization objective. This can lead to the inconsistency between the construction and the use of demand curves, and the sub-optimality of the market clearing results. In the following two sections, a sequential clearing algorithm and an iterative clearing algorithm are analyzed and shown to produce socially suboptimal clearing results.

III. Sequential Clearing of Multiple Products with Demand Curves

In this section, we first describe a sequential clearing process implemented in some existing markets under the simplified two-product setting; then we analyze why it fails to produce the socially optimal solution; lastly, we discuss the implied interpretation of demand curves from the sequential process, and present the corresponding social surplus maximization solution.

III.1 A sequential clearing process for the two-product market

Consider the simple two-product setting of the previous section with a high-quality product *A* in zone *A* and a low-quality product *B* in the rest-of-system zone *B*. The topology of the system is described in Fig. 6.

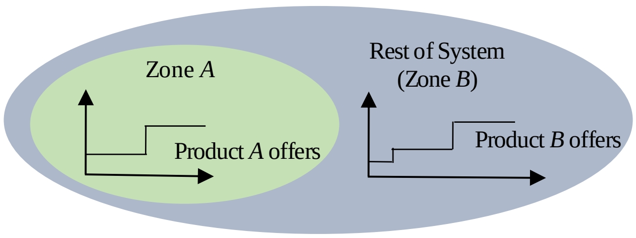

Fig.6. A two-zone system with two products.

The aggregated offers for products or zones[2] *A* and *B*, respectively, are illustrated in the above Fig.6. A local demand curve for zone *A*, and a system demand curve for the entire system (including both zones) are used in the clearing process. Note that the high-quality product in zone *A* is counted toward the system demand while the low-quality product *B* cannot count toward the demand for product *A*.

In a typical sequential clearing process, the market sequentially clears the system and zonal products by intersecting the aggregated supply offers with the corresponding demand curves. The process is described as follows:

Step 1. Find the intersection of the system demand curve and the aggregated system supply curve based on offers from both zones. Calculate the price λ_{SYS} of the intersection point.

Step 2. Find the intersection of the local demand curve in zone *A* and the aggregated supply offers from zone *A*. Calculate the price λ_A of the intersection point.

Step 3. Depending on the relation between λ_A and λ_{SYS}, two possible clearing results follow:

3a. If $\lambda_A < \lambda_{SYS}$, then λ_{SYS} will be the market clearing price for both zones/products, and all the supply offers that are to the left of the intersection in *Step 1* will be cleared.

3b. Otherwise if $\lambda_A \geq \lambda_{SYS}$, then λ_A will be the market clearing price for product *A*, and all product *A* offers that are to the left of the intersection in *Step 2* will be cleared. A new aggregated system offer curve obtained by shifting the product *B*'s aggregated supply offers to the right by the cleared quantity of product *A* (i.e., treating product *A*'s cleared quantity as zero price) will be formed to intersect with the system demand curve. Or equivalently, *the system demand curve is shifted to the left by the cleared quantity of product A* to intersect with the aggregated offers for product *B*. The price λ_B associated with the intersection will be the market clearing price for zone *B*, and all product *B* offers that are to the left of the intersection will be cleared.

Note that the above algorithm applies to both fixed requirements (consider them as vertical demand curves) and sloped demand curves. The above clearing process is somewhat "*intuitive*": When the high-quality product *A* is "inexpensive" (i.e., $\lambda_A < \lambda_{SYS}$ under *Step 3a*), there's no price separation between the two zones, and the above process clears the market as if there is only one system-wide low-quality product *B* in the market, which leads to the *Step 1* solution; When product *A* is "expensive" (i.e., $\lambda_A \geq \lambda_{SYS}$ under *Step 3b*), there is price separation between the two zones. Product *A* will clear at the intersection of the aggregated supply curve for *A* and zone *A*'s demand curve, and subsequently product *B* will clear at the intersection of the aggregated supply curve for *B* and zone *B*'s residual demand curve (which is obtained by shifting the system demand curve to the left by the cleared amount of product *A*).

[2] In a zonal market, each zone defines a product. Therefore, the terms "zone" and "product" are used interchangeably hereafter in this paper.

The above intuitive solution can be proven optimal in terms of production cost minimization with fixed requirements, assuming the problem is feasible (See proof in *Appendix A*). However, when demand curves are adopted, demand becomes a variable whose benefits need to be formulated and included in the market clearing objective. The following analysis shows that the demand benefit is not appropriately represented in the above sequential clearing process.

III.2 Issues with the sequential clearing process

The issue with the sequential clearing process described in the previous subsection III.1 can be demonstrated with the price separation case in *Step 3b*, where the two market products are sequentially cleared at the intersection points of corresponding demand curves and supply curves. Such clearing result is illustrated in Fig. 7.

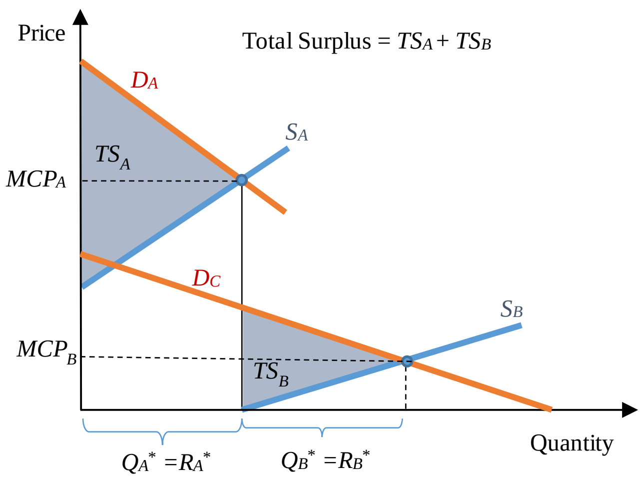

Fig.7. Sequential clearing of two products with price separation.

In the above figure, the high-quality product *A* is cleared first at the intersection of *A*'s demand curve D_A and aggregated supply curve S_A, resulting in the cleared supply quantity Q_A^* (equal to the cleared demand R_A^*) and the market clearing price MCP_A for product *A*. Then product *B*'s aggregated supply curve S_B is shifted to the right by Q_A^* to intersect with the system demand curve D_C. Or equivalently, the system demand curve D_C is shifted to the left by R_A^* to yield a residual demand curve for product *B*, i.e., $D_C(R + R_A^*)$, to intersect with *B*'s aggregated supply curve S_B, resulting in the cleared supply quantity Q_B^* (equal to the cleared demand R_B^*) and the market clearing price MCP_B for product *B*. In the above sequential clearing of *A* and *B*, the social surplus associated with each product is maximized in their corresponding clearing steps. Namely, the shaded area TS_A represents the maximal social surplus for product *A*, and the shaded area TS_B represents the maximal social surplus for product *B* given *A*'s cleared quantity.

However, the market clearing problem should maximize the total surplus of the two products, i.e., $TS_A + TS_B$. By fixing product *A*'s quantity at Q_A^* or R_A^* in deriving *B*'s clearing results, the above sequential process fails to capture the impact of *A*'s cleared quantity on the demand benefit of *B*. For instance, if the market clears a little less than the supply quantity Q_A^* and the demand quantity R_A^* for product *A*, then the impact on the social surplus associated with *A* (i.e., the shaded area TS_A) is ignorable but the residual demand curve for *B* will be higher, resulting in a marginal gain of $D_C(R_A^*) - D_C(R_A^* + R_B^*)$ for the social surplus associated with *B* (i.e., the shaded area TS_B) and the total social surplus. Therefore, the sequential clearing process leads to a suboptimal solution in the above price-separation situation.

III.3 Social surplus maximization

As discussed in the above subsection, the sequential clearing process may not yield the overall optimal results because it did not appropriately account for the impact of product *A*'s clearing on the demand benefit of product *B*. In fact, the sequential clearing process does not present a clear form of the total social surplus. To address the issue, the form of total demand benefits, which relies on the interpretation of demand curves, must be specified. As discussed in subsection II.4, given the demand curve D_A for the high-quality product *A* and the system demand curve D_C, additional assumptions are needed in order to evaluate the overlapped demand benefit B_{AC} in Fig. 5 and Eq. (32). In the sequential *Step 3b* described in subsection III.1, the system demand curve D_C is shifted by product *A*'s cleared quantity, which suggests that the product A's benefit under the system demand curve D_C is considered to be associated with the high-price portion, i.e.,

$$B_{AC} = \int_0^{R_A} D_C(R)\, dR. \quad (38)$$

This is the same assumption (34) discussed in subsection II.4. Therefore, the total benefit function (32) is represented as:

$$Total\ Benefit = \int_0^{R_A} D_A(R)\, dR + \int_0^{R_C} D_C(R)\, dR - \int_0^{R_A} D_C(R)\, dR. \quad (39)$$

Combining the first and the last terms in the above benefit expression, we have

$$Total\ Benefit = \int_0^{R_A} (D_A(R) - D_C(R))\, dR + \int_0^{R_C} D_C(R)\, dR. \quad (40)$$

Define $D_A'(R)$ as the additional marginal benefit of product *A* over product *B*, i.e.,

$$D_A'(R) \equiv D_A(R) - D_C(R). \quad (41)$$

Then the social surplus maximization problem is represented by (18)-(22) in subsection II.2. And D_A' and D_C have the demand curve interpretations in subsection II.2. Solving the optimization problem (18)-(22) leads to the maximal social surplus as will be demonstrated by the numerical examples in Section V. The optimization problem (18)-(22) can be solved by existing solvers such as CPLEX, or by modifying the current sequential clearing process[3].

Note that the demand curve D_A' derived from (36) may not be monotonically decreasing, even if both the curve D_A and D_C are monotonically decreasing. This implies that the social surplus objective (18) may be nonconvex with D_A' derived from (36). Solving the non-convex market clearing problem is out of the scope of this paper.

[3] The modification of the existing sequential process to achieve the social optimal solution will not be discussed in this paper, as directly solving the optimization problem (18)-(22) with solvers presents a straightforward solution.

IV. ITERATIVE CLEARING OF MULTIPLE PRODUCTS WITH DEMAND CURVES

In this section, we first describe an iterative clearing process adopted in some existing markets using the two-product setting; then we analyze the convergence of the iterative algorithm and the sub-optimality of the iterative solution; finally, the proper form of social surplus optimization is discussed.

IV.1 An iterative clearing process for the two-product market

Consider the two-zone system depicted in Fig. 6. Denote the total system by zone *C* (i.e., $\mathrm{C} = \mathrm{A} \cup \mathrm{B}$). Two demand curves are given prior to the market clearing: the demand curve D_A for zone *A*, and the demand curve D_C for the system or zone *C*. Market participants submit supply offers for the high-quality product in zone *A* or the low-quality product in zone *B*. The product in zone *A* can be used to meet the demand in the system zone *C*.

With the two demand curves and product offers in zone *A* and zone *B*, the iterative clearing process solves a series of optimization problems. Each optimization problem has the similar mathematical formulation of minimizing the total costs of cleared product offers minus the integrals of two *updated* demand curves for the local zone *A* and the rest of system zone *B*, subject to zone *A*'s product requirement, zone *C*'s product requirement and other individual offer constraints. The formulation for the *k*-th iteration is represented as:

$$\min_{Q_A,Q_B,R_A,R_B} C_A(Q_A) + C_B(Q_B) - \int_0^{R_A} D_A(R)\,dR \quad - \int_{R_A^{k-1}}^{R_A^{k-1}+R_B} D_C(R)\,dR \qquad (42)$$

$$s.t. \qquad Q_A \geq R_A \qquad (43)$$

$$Q_A + Q_B \geq R_A + R_B \qquad (44)$$

$$Q_A \in \Omega_A,\ Q_B \in \Omega_B \qquad (45)$$

$$R_A, R_B \geq 0. \qquad (46)$$

$$R_A \leq R_A^{k-1}. \qquad (47)$$

where

Q_A, Q_B: Cleared supply offers of zones *A* and *B*, respectively;

$C_A(\cdot), C_B(\cdot)$: Supply offer costs of zones *A* and *B*, respectively;

Ω_A, Ω_B: Product constraint sets of products *A* and *B*, respectively;

R_A, R_B: Cleared demand in zones *A* and *B*, respectively;

R_A^{k-1}: Cleared demand of the zone *A* from the (*k*-1)-th iteration.

In the above formulation, the term $\int_{R_A^{k-1}}^{R_A^{k-1}+R_B} D_C(R)\,dR$ in the objective function reflects the shifted demand curve $D_C(\cdot)$ by R_A^{k-1} of the previous iteration, and constraint (47) is equivalent to truncating the demand curve $D_A(\cdot)$ at R_A^{k-1} (i.e., there would be no benefit of clearing more than R_A^{k-1}). The shifting and truncating of the original demand curves are illustrated in Fig. 8.

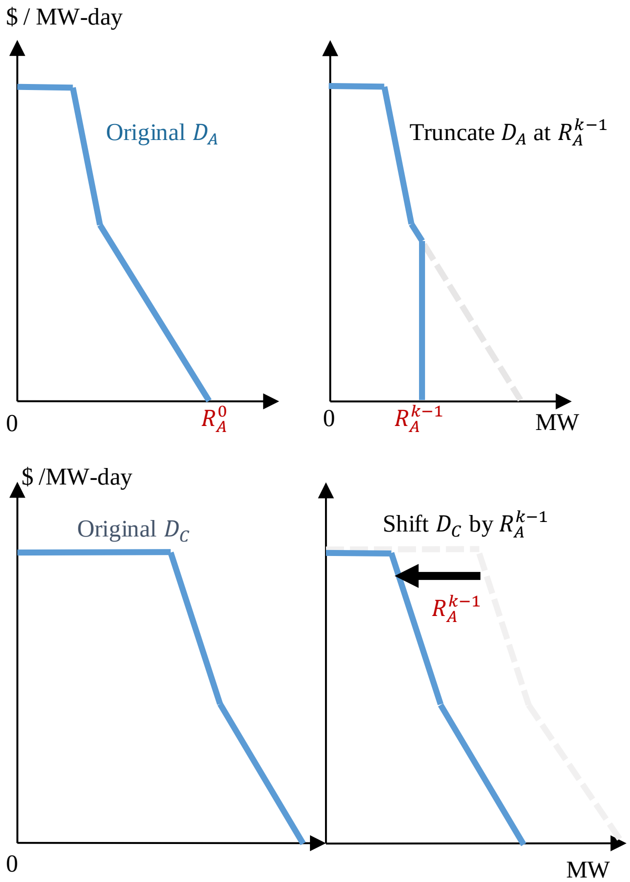

Fig. 8. Truncating of zone A's demand curve and shifting of system zone C's demand curve.

The iterative process is described as follows:

Iteration 0 (Initialization). Find the maximal cleared demand in zone A at where the zone's demand curve descends to zero price, i.e., $R_A^0 = \min\{R_A: D_A(R_A) \leq 0\}$.

Iteration k. Truncate zone A's original demand curve D_A at the zone's cleared demand R_A^{k-1} from the previous iteration, and shift system zone C's original demand curve D_C to the left by R_A^{k-1} (illustrated by Fig.8). Then solve the optimization problem with the updated demand curves, i.e., (42)-(47), to obtain R_A^k.

Stopping criteria. The truncating of child zone's demand curve, the shifting of parent zone's demand curve and the optimization will repeat until the cleared demand quantity for zone A is equal to that in the previous iteration, i.e., $R_A^k = R_A^{k-1}$. Clearing prices are obtained from the last iteration.

IV.2 Sub-optimality of the iterative clearing process

First, we prove the convergence of the above iterative clearing process in the following theorem:

Theorem 2. The iterative process described in subsection IV.1 converges to a solution.

[*Proof*]. With the truncating of zone A's demand curve $D_A(\cdot)$, or equivalently with constraint (47) in each iteration, the cleared demand in zone A decreases as iterations increase, i.e., $R_A^k \geq R_A^{k+1}$. Since the cleared demand in zone A in any iteration is also lower bounded by zero with constraint (46), the monotonic sequence of iterative solutions, i.e., $\{R_A^k\}_{k=1,2,\cdots}$, must converge to the infimum of the sequence. [*End of Proof*].

In the above iterative process, the optimization problem (42)-(47) solved in each iteration resembles the clearing model (7)-(11) of subsection II.1 with the term $\int_{R_A^{k-1}}^{R_A^{k-1}+R_B} D_C(R)\,dR$ being the k-th iteration demand benefit for zone

B. The demand curve $D_B^k(\cdot)$ of zone *B* (i.e., the derivative of the above term), is updated in the *k*-th iteration by shifting zone *C*'s demand curve $D_C(\cdot)$ by R_A^{k-1} from the previous iteration, i.e., $D_B^k(R) = D_C\left(R + R_A^{k-1}\right)$.

Suppose that the cleared demand in zone *A* converges to R_A^* in the iterative clearing process. Then R_A^* must be the solution to the following problem:

$$\min_{Q_A,Q_B,R_A,R_B} C_A(Q_A) + C_B(Q_B) - \int_0^{R_A} D_A(R)\,dR \;\; - \int_{R_A^*}^{R_A^*+R_B} D_C(R)\,dR \qquad (48)$$

s.t. (43)-(46) and $R_A \leq R_A^*$.

Denote the total benefit (TB) of the two products in (48) by *TB**, i.e.,

$$TB^* = \int_0^{R_A} D_A(R)\,dR + \int_{R_A^*}^{R_A^*+R_B} D_C(R)\,dR \qquad (49)$$

Then at the convergence, the marginal benefit of zone *A*'s demand at R_A^* is

$$MB^*(R_A^*) = \left.\frac{\partial TB^*}{\partial R_A}\right|_{R_A=R_A^*} = D_A(R_A^*) \qquad (50)$$

To show the sub-optimality of the iterative solution R_A^*, now consider optimizing R_A and R_B in a single optimization problem (without the iterative update). Instead of (49), the total demand benefit (TB) of the two products is:

$$TB = \int_0^{R_A} D_A(R)\,dR + \int_{R_A}^{R_A+R_B} D_C(R)\,dR. \qquad (51)$$

Denote the optimal demands of zone *A* and zone *B* under the single optimization as R_A^{opt} and R_B^{opt}, respectively. Therefore the marginal benefit of zone *A*'s demand at R_A^{opt} is

$$MB^{opt}\left(R_A^{opt}\right) = \left.\frac{\partial TB}{\partial R_A}\right|_{R_A=R_A^{opt}} = D_A\left(R_A^{opt}\right) - D_C\left(R_A^{opt}\right) + D_C\left(R_A^{opt} + R_B^{opt}\right). \qquad (52)$$

Unlike the marginal benefit (50) associated with the iterative solution R_A^*, it can be seen that the marginal benefit (52) under the single optimization takes into account the marginal impact of zone *A*'s demand on zone *B*'s benefit of $\int_{R_A}^{R_A+R_B} D_C(R)\,dR$, i.e.,

$$\partial \int_{R_A}^{R_A+R_B} D_C(R)\,dR \big/ \partial R_A = -D_C(R_A) + D_C(R_A + R_B). \qquad (53)$$

Note that $D_C(R_A + R_B) \leq D_C(R_A)$ since the system demand curve D_C is monotonically decreasing. So increasing zone *A*'s demand has an impact of reducing zone *B*'s benefit. By ignoring such impact, the iterative solution overvalues the marginal benefit of zone *A*'s demand, which leads to the over-clearing of the high-quality product in zone *A*. In sum, the currently implemented iterative process may not produce the social optimal solution. The sub-optimality of the iterative process will be further demonstrated by numerical examples in Section V.

IV.3 The optimal clearing problem

In the iterative process described in subsection IV.1, the iterative objective function (42) implies that the intended social surplus optimization problem is

$$\min_{Q_A,Q_B,R_A,R_B} C_A(Q_A) + C_B(Q_B) - \int_0^{R_A} D_A(R)\,dR - \int_{R_A}^{R_A+R_B} D_C(R)\,dR \qquad (54)$$

s.t. (43)-(46).

Note that the total demand benefit in (54) can be reorganized as follows:

$$\int_0^{R_A} D_A(R)\,dR + \int_{R_A}^{R_A+R_B} D_C(R)\,dR$$
$$= \int_0^{R_A} (D_A(R) - D_C(R))\,dR + \int_0^{R_A+R_B} D_C(R)\,dR \qquad (55)$$

Denote $R_C \equiv R_A + R_B$ and $D_A'(R) \equiv D_A(R) - D_C(R)$. Then the demand benefit can be further represented by

$$TB = \int_0^{R_A} D_A'(R)\,dR + \int_0^{R_C} D_C(R)\,dR. \qquad (56)$$

Thus the demand curve interpretations for $D'_A(\cdot)$ and $D_C(\cdot)$ fall into those of subsection II.2. Solving the optimization problem (18)-(22) by existing solvers leads to the maximal social surplus. Note that the demand curve D'_A derived from $D_A - D_C$ may not be monotonically decreasing even if both D_A and D_C are monotonically decreasing. Therefore, the objective function of the social optimization problem may not be convex. Issues related to solving a non-convex market clearing problem are out of the scope of this paper.

V. NUMERICAL EXAMPLES

In this section, we use a two-product capacity market to illustrate the solution sub-optimality of the sequential and iterative clearing processes in Section III and Section IV by comparing their results to the socially optimal solution. Two examples are presented, with Example 1 demonstrating the potential sub-optimality of the quantity allocation and Example 2 demonstrating the potential sub-optimality of the clearing prices from the sequential or iterative clearing process.

Example 1. Consider a two-zone capacity market. Two capacity products, zone A capacity and zone B (rest of system) capacity, will be cleared in the market with the requirements for zone A and the system zone C ($C = A \cup B$). Capacity in zone A (e.g., an import-constrained zone) has a higher quality as it can be used to meet both the zone A and the system capacity requirements. Capacity in zone B (e.g., the rest of system) has a lower quality as it can only be used to meet the system capacity requirement. Capacity supply offers in the two zones are listed in Table 1 and illustrated in Fig. 9. The demand curve parameters are listed in Table 2 and illustrated in Fig. 10. All capacity quantities are in GW, and all prices are in $/kw-month. Three cases are examined: Case 1 with the existing sequential clearing process, Case 2 with the existing iterative clearing process, and Case 3 with the social surplus maximization.

Table 1. Capacity offers for Example 1

Quantity (GW)	Price ($/kw·month)	Location
1	5	Zone *A*
2	1	Zone *B*

Table 2. Demand curves for Example 1

	Demand Curves ($/kw·month)
Zone *A*	10×(1-*R*)
System	5×(1-*R*/2)

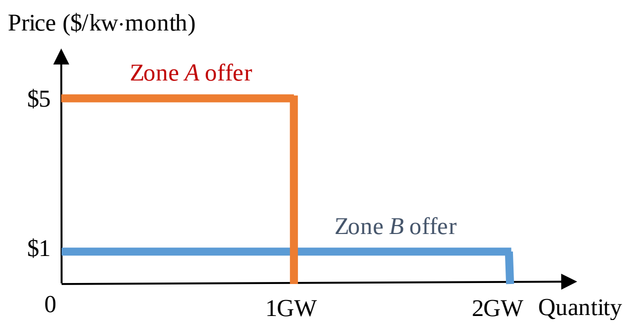

Fig. 9. Zonal capacity offers for Example 1.

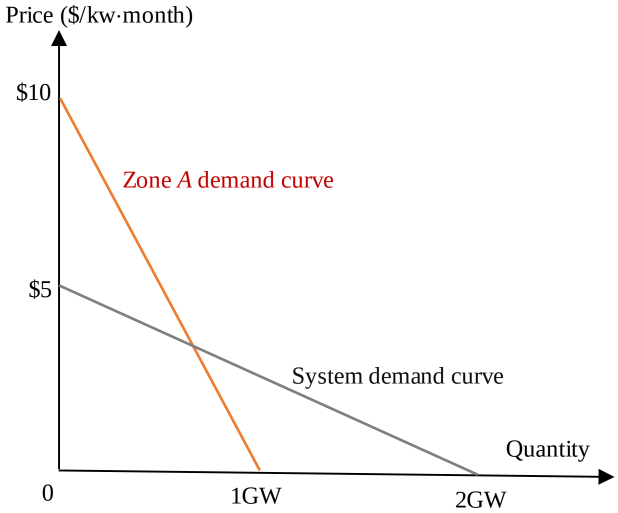

Fig. 10. Demand curves for Example 1.

Case 1: Sequential clearing process.

Consider first the sequential clearing steps described in subsection III.1. *Step 1* results in the intersection of all offers and the system demand curve at the price of $1/kw-month. *Step 2* results in the intersection of zone *A*'s supply offers and product *A*'s demand curve at the price of $5/kw-month. Since the *Step 2* price is higher than the *Step 1* price, *Step 3b* follows. The cleared 0.5GW capacity in zone *A* from *Step 2* will be treated as zero-price offer, which is combined with zone *B*'s offers to form an aggregated offer curve to intersect with the system demand curve. The price at the intersection is $1/kw-month, and the system capacity is cleared at 1.6GW, implying that zone *B*'s capacity is cleared at 1.1GW (=1.6GW–0.5GW). This sequential clearing process is illustrated in Fig. 11 and the clearing results are summarized in Table 3.

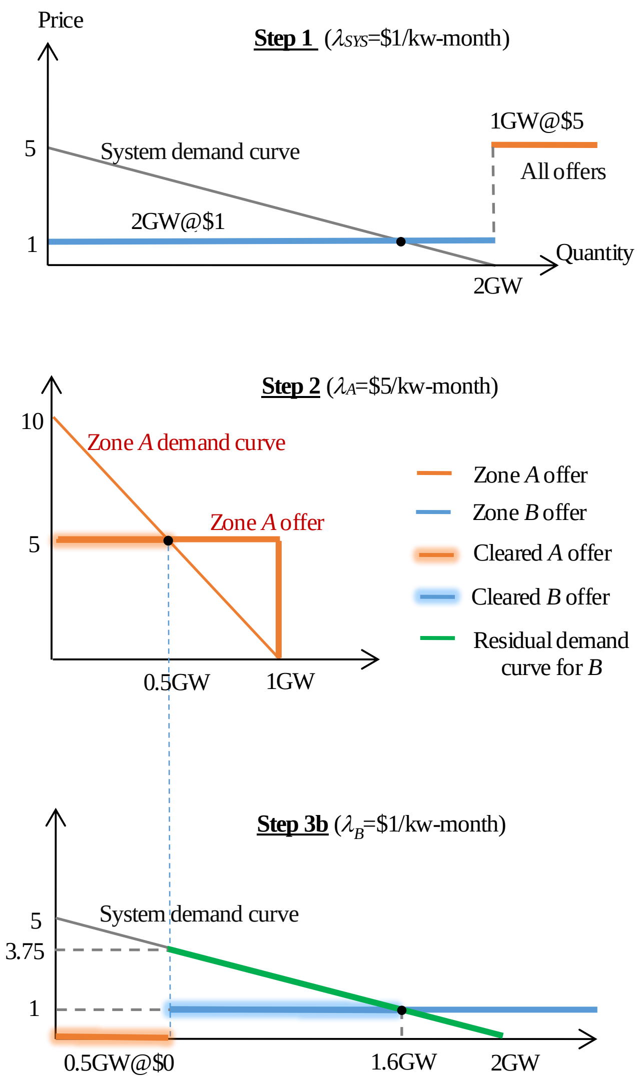

Fig.11. The sequential clearing process for Example 1.

Table 3. Sequential clearing results for Example 1

	Cleared Capacity (GW)	Cleared Demand (GW)	Clearing Price ($/kw-month)
Zone *A*	0.5	0.5	5
Zone *B*	1.1	1.1	1
Total Surplus = $\$2.7625\times10^6$ /month			

To check whether the above result maximizes the social surplus, we consider a small reduction of Δ kw to the cleared capacity in zone *A*. The capacity cost in zone *A* is then reduced by \$5Δ. Since the cleared capacity in zone *A* must be no less than the cleared demand in the zone, zone *A*'s demand is reduced by Δ kw, and the zone *A*'s demand benefit is reduced by \$5Δ. Therefore, the net change of the social surplus associated with product *A* is 0. Now consider the impact of reducing zone *A*'s cleared capacity on the social surplus associated with product *B*. In *Step 3b* illustrated in the above Fig. 11, it can be seen that reducing *A*'s cleared capacity would raise the residual demand curve for zone *B* and thus increases zone *B*'s cleared demand by Δ and its demand benefit by \$3.75Δ. Also, zone *B*'s cleared capacity will increase by Δ kw to cover the increase of demand in *B*, with an increased cost of \$Δ. As a result, the total surplus of the two products is increased by \$2.75Δ (= \$3.75Δ – \$Δ), and thus the current sequential clearing solution is *not* optimal. The issue is that *Step 3b* of the sequential process ignores the marginal impact of increasing zone *B*'s demand benefit by reducing zone *A*'s capacity, causing over-purchase of zone *A*'s capacity and under-purchase of zone *B*'s capacity. It can be calculated that the social surplus associated with the sequential solution is $2.7625×10^6 /month.

Case 2: Iterative clearing process.

Based on the description of the iterative algorithm in subsection IV.1, the initial value for the cleared demand in zone *A* is obtained at Q_A^0 = 1GW (where the zone *A*'s demand curve reaches zero price). The iterative optimization problems are then solved by using a CPLEX optimization solver. The iterative results until convergence are listed in Table 4.

Table 4. Iterative solutions for Example 1

Iteration k	**Cleared Capacity (MW)** (Q_A^k, Q_B^k)	**Cleared Demand (MW)** (R_A^k, R_B^k)	**Clearing Prices ($/kw-month)** (MCP_A^k, MCP_B^k)	**Social Surplus ($/month)** (obj^k)
0	(*, *)	(**1.0**, *)	(*, *)	*
1	(0.5, 0.6)	(**0.5**, 0.6)	(5, 1)	$1.700×10^6$
2	(0.5, 1.1)	(**0.5**, 1.1)	(5, 1)	$\mathbf{2.7625×10^6}$

It can be seen from the above table that the iterative algorithm converged within three iterations. At convergence, the capacity and demand cleared in zone *A* is 0.5 GW; the capacity and demand cleared in zone *B* is 1.1 GW (the total demand cleared in the system is 1.6GW); and the clearing prices for the two zones are $5/kw-month and $1/kw-month, respectively. The social surplus at the convergence of the iterative process is $2.7625×10^6 /month.

Note that the above iterative solution in Table 4 is the same as the sequential solution in Case 1. This is not a coincidence. A discussion on the connection between the sequential and the iterative clearing processes of Section III and Section IV is out of the scope of this paper.

Case 3: Socially optimal solution.

As discussed in subsections III.3 and IV.3, both the sequential and the iterative processes imply the assumption that the basic demand benefit of the high-quality product *A* is associated with the high-value part of the system demand curve, i.e., (34). Under the assumption, one can derive zone *A*'s additional marginal benefit curve D_A' from *A*'s demand curve D_A and the system demand curve D_C. Demand curves D_A' and D_C have the interpretations described in subsection II.2. The resulting social surplus optimization problem is (18)-(22).

For the demand curves in Table 2 of this example, the additional marginal benefit for *A* would be

$$D_A'(R) = D_A(R) - D_C(R) = 10 \times (1 - R) - 5 \times (1 - R/2) = 5 \times (1 - 3R/2).$$

Then solving the market clearing problem (18)-(22) using CPLEX yields the below optimal solution in Table 5.

Table 5. Social surplus optimization solution for Example 1

	Cleared Capacity (GW) (Q_A^{opt}, Q_B^{opt})	**Cleared Demand (GW)** (R_A^{opt}, R_C^{opt})	**Clearing Prices $/kw-month** $(MCP_A^{opt}, MCP_B^{opt})$	**Social Surplus ($/month)** (obj^{opt})
Optimal Solution	(0.133, 1.467)	(**0.133**, 1.6)	(5, 1)	$3.2667×10^6$

Compared to the sequential solution in Case 1 or the iterative solution in Case 2, the market clearing prices and the total cleared system capacity under the social optimal solution happen to be unchanged. However, the allocation of cleared capacity between zone *A* and zone *B* are quite different between the optimal solution and the sequential/iterative solution, e.g., the optimally cleared capacity in zone *A* is only 0.133GW compared to 0.5GW under the suboptimal solution. The social surplus optimization solution allocates less capacity for zone *A* (and more capacity for the rest-of-system zone *B*) since it appropriately accounts for zone *B*'s benefit increase as a result of reducing zone *A*'s capacity. The additional gain of social surplus under the optimal solution is $0.5042×10^6 /month, or an 18.3% increase from the suboptimal surplus under the sequential/iterative solution.

Example 2. Consider the same settings as Example 1 except that the capacity offers for the two products are replaced with the ones listed in Table 6 and illustrated in Fig. 12.

Table 6. Capacity offers for Example 2

Quantity (GW)	Price ($/kw·month)	Location
1	$1\times Q_A$	Zone A
2	$0.8\times Q_A$	Zone B

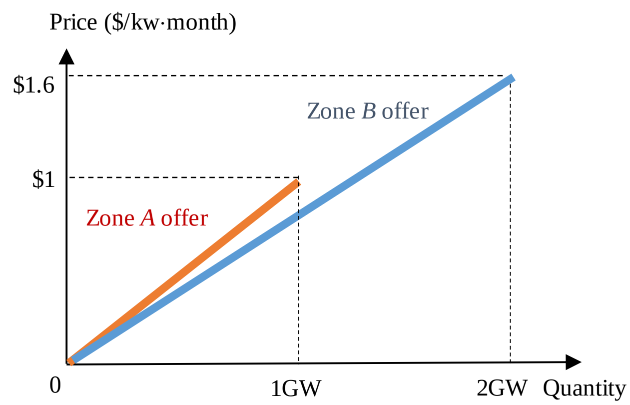

Fig. 12. Zonal capacity offers for Example 2.

Case 1: Sequential clearing process.

The sequential clearing process for this example is illustrated in Fig. 11.

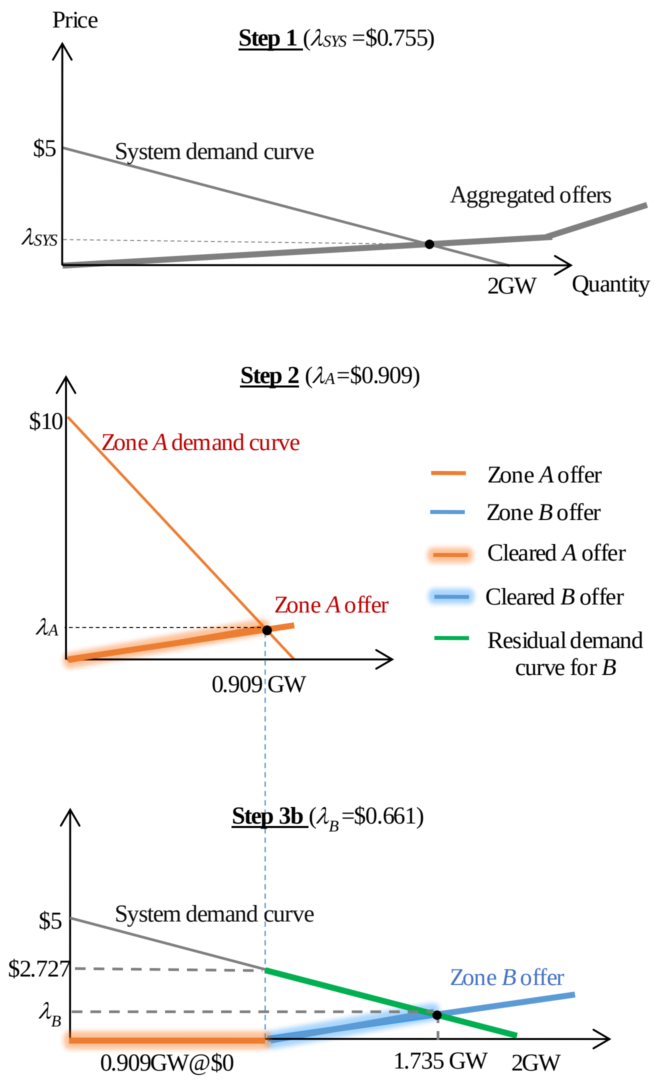

Fig.13. The sequential clearing process for Example 2.

Table 7. Sequential clearing results for Example 2

	Cleared Capacity (GW)	Cleared Demand (GW)	Clearing Price ($/kw-month)
Zone *A*	0.909	0.909	0.909
Zone *B*	0.826	0.826	0.661
Total Surplus = $**5.6724×10^6**/month			

In the sequential clearing process illustrated in the above Fig.13, *Step 1* results in the intersection of aggregated offers of both zones and the system demand curve at the price of $0.755/kw-month. *Step 2* results in the intersection of zone *A*'s supply offers and product *A*'s demand curve at the price of $0.909/kw-month. Since the *Step 2* price is higher than the *Step 1* price, *Step 3b* follows. The cleared 0.909GW capacity in zone *A* from *Step 2* will be treated as zero-price offer, which is combined with zone *B*'s offers to form an aggregated offer curve to intersect with the system demand curve. The price at the intersection is $0.661/kw-month, and the system capacity is cleared at 1.735GW, implying that zone *B*'s capacity is cleared at 0.826GW (=1.735GW–0.909GW). The clearing results are summarized in Table 7. It can be seen that the sequential clearing process leads to the price separation between zone *A* and zone *B*.

Case 2: Iterative clearing process.

With the iterative algorithm in Section IV.1 applied to this example 2, the initial value for the cleared demand in zone A is obtained at Q_A^0 = 1GW (where the zone A's demand curve reaches zero price) and the iterative results are listed in Table 8.

Table 8. Iterative solutions for Example 2

Iteration $\boldsymbol{k}$	**Cleared Capacity (GW)** (Q_A^k, Q_B^k)	**Cleared Demand (GW)** (R_A^k, R_B^k)	**Clearing Prices ($/kw-month)** (MCP_A^k, MCP_B^k)	**Social Surplus ($/month)** (obj^k)
0	(*, *)	(**1.0**, *)	(*, *)	*
1	(0.909, 0.758)	(**0.909**, 0.758)	(0.909, 0.606)	5.4924×10^6
2	(0.909, 0.826)	(**0.909**, 0.826)	(0.909, 0.661)	$\mathbf{5.6724\times10^6}$

It can be seen from the above table that the iterative algorithm converged within three iterations. At convergence, the capacity and demand cleared in zone A is 0.909 GW; the capacity and demand cleared in zone B is 0.826 GW (the total demand cleared in the system is 1.735GW); and the clearing prices for the two zones are $0.909/kw-month and $0.661/kw-month, respectively. Note that the iterative process leads to price separation between zone A and zone B. The social surplus at the convergence of the iterative is $\$5.6724\times10^6$ /month. The solution is the same as the sequential solution in Table 7 of Case 1.

Case 3: Socially optimal solution.

As discussed in subsections III.3 and IV.3, both the sequential and the iterative processes imply the assumption of (34). Under the assumption, one can derive zone A's additional marginal benefit curve D_A' from A's demand curve D_A and the system demand curve D_C. Demand curves D_A' and D_C have the interpretations described in subsection II.2. The resulting social surplus optimization problem is (18)-(22).

For the demand curves in Table 2 of this example, the additional marginal benefit curve for A would be

$$D_A'(R) = D_A(R) - D_C(R) = 10 \times (1 - R) - 5 \times (1 - R/2) = 5 \times (1 - 3R/2).$$

Then solving the market clearing problem (18)-(22) using CPLEX yields the below optimal solution in Table 9.

Table 9. Social surplus optimization solution for Example 2

	Cleared Capacity (GW) (Q_A^{opt}, Q_B^{opt})	**Cleared Demand (GW)** (R_A^{opt}, R_C^{opt})	**Clearing Prices ($/kw-month)** $(MCP_A^{opt}, MCP_B^{opt})$	**Social Surplus ($/month)** (obj^{opt})
Optimal Solution	(0.755, 0.943)	(0.667, 1.698)	(0.755, 0.755)	$\mathbf{5.9119\times10^6}$

Compared to the price separation result of the sequential solution in Case 1 or the iterative solution in Case 2, the above optimal solution yields no price separation between product A and product B. The optimal clearing price for product A is (0.909-0.755)/0.755=20.4% lower than its sequential/iterative clearing price, while the optimal clearing price for product B is (0.755-0.661)/0.755=13.8% higher than its sequential/iterative clearing price. Compared to the socially optimal solution, the sequential/iterative process over cleared 0.154GW (=0.909-0.755) or 20.4% (=0.154/0.755) of capacity in zone A, and under cleared 0.117GW (=0.943-0.826) or 12.4% (=0.117/0.943) of capacity in zone B.

The above two examples show a significant optimality gap between the sequential/iterative solution and the optimal social surplus solution. The results also reveal potentially large impacts on individual capacity resources as a result of the disparity in capacity allocation among the zones under different solutions. Considering that the clearing prices are also likely to be different under the optimal and sequential/iterative solutions (as demonstrated in Example 2), the combined impact could be even bigger. In an era with decreasing marginal energy costs, generators would increasingly rely on the capacity market (if one is in place). Therefore, the sub-optimality issue of the sequential or iterative clearing in existing capacity markets is worthy of a close examination.

VI. CONCLUSION

Demand curves are often treated as a simple upgrade to fixed product requirements and penalty factors in ISO markets. Therefore, in transition to demand curves, the interpretation of demand curves is rarely discussed and the existing market clearing procedures are presumed to hold. However, with demand curves, the market clearing objective is no longer a cost minimization but social surplus optimization that includes demand-side benefit. Consequently, a proper economic interpretation of demand curves and the subsequent composition of demand benefit are fundamental to forming the social surplus, in particular when there are multiple coupled products in the market. Using a two-product market setting, we reveal two different interpretations of the demand curves, each associated with a specific form of the clearing problem formulation. Our findings will ensure consistency between the construction of demand curves and their use in the market clearing. Also, we demonstrate that the current sequential and iterative market clearing processes used in some ISOs may not produce socially optimal solutions when multi-product demand curves are involved. The cause of this is attributed to the lack of clear interpretation of demand curves and demand benefits. The proper interpretation of the demand curves in these markets is proposed to form the social surplus optimization problem. Our work provides a much needed guide for the proper construction and use of demand curves at a time when the demand curve is being increasingly used as an important pricing tool for the markets with significant penetration of near-zero marginal cost resources.

REFERENCES

[1] W.W. Hogan, "Electricity Scarcity Pricing Through Operating Reserves," April 2013. [Available:] https://sites.hks.harvard.edu/fs/whogan/

[2] ERCOT, "ORDC Workshop," (Training material), [Available:] http://www.ercot.com/content/wcm/training_courses/107/ordc_workshop.pdf

[3] NYISO, Manual 4: Installed Capacity Manual. September 2019. [Available:] https://www.nyiso.com/manuals-tech-bulletins-user-guides

[4] PJM, Manual 18: PJM Capacity Market. December 2019. [Available:] https://www.pjm.com/directory/manuals/m18/index.html

[5] PJM, "Enhanced Price Formation in Reserve Markets of PJM Interconnection, L.L.C," FERC Docket No. EL19-58-000. [Available:] https://pjm.com/directory/etariff/FercDockets/4036/20190329-el19-58-000.pdf

[6] ISO-NE, Market Rule 1, Section 13.2. [Available:] https://www.iso-ne.com/static-assets/documents/regulatory/tariff/sect_3/mr1_sec_13_14.pdf

[7] F. Zhao, T. Zheng, E. Litvinov, "Constructing Demand Curves in Forward Capacity Market," *IEEE Transactions on Power Systems*, March 2017, *Vol*.33, *No*. 1, pp. 525-535.

APPENDIX A: OPTIMALITY OF THE SEQUENTIAL CLEARING PROCESS WITH FIXED REQUIREMENTS AND FEASIBILITY ASSUMPTION

In this Appendix, we prove that the sequential clearing steps described in subsection II.1 will lead to the same results as the cost minimization problem with fixed product requirements, *if the problem is feasible*. Consider that the fixed requirements for product A and the system are R_A and R_C, respectively. The optimization problem for the market clearing with fixed requirements can be represented as:

$$\min_{Q_A,Q_B} C_A(Q_A) + C_B(Q_B) \qquad \text{(A1)}$$

$$s.t. \quad Q_A \geq R_A, \qquad (\lambda'_A) \qquad \text{(A2)}$$

$$Q_A + Q_B \geq R_C, \qquad (\lambda_C) \qquad \text{(A3)}$$

$$Q_A \in \Omega_A,\ Q_B \in \Omega_B. \qquad \text{(A4)}$$

where the objective (A1) is to minimize the total production cost in the absence of demand benefit measurements. *Suppose that the above optimization problem is convex*, and the shadow prices associated with (A2) and (A3) are λ'_A and λ_C, respectively. Then the clearing price for the high-quality product *A* is $(\lambda'_A + \lambda_C)$, and the clearing price for the low-quality product *B* is λ_C. The following theorem holds:

***Theorem A*. With fixed requirements, the sequential clearing steps lead to the same cleared quantities and clearing prices as the optimization problem (A1)-(A4) if the problem is feasible**.

[*Proof*]. *Step 1* of the sequential clearing process is equivalent to solving an optimization problem (A1) and (A3)-(A4) without the zone *A* requirement constraint (A2), and the shadow price associated with (A3) is the intersection price λ_{SYS}. *Step 2* is equivalent to minimizing the cost of meeting zonal requirement R_A without considering the system requirement, i.e., $\min_{Q_A} C_A(Q_A)$ subject to (A2), and the shadow price associated with (A2) is the intersection price λ_A. Since the problem is assumed feasible, then both intersections exist. Now consider two cases:

- $\lambda_A<\lambda_{SYS}$. This means product *A*'s cleared quantity in *Step 2* is less than its quantity cleared in *Step 1*. As a result, solving (A1)-(A4) is equivalent to the *Step 1* optimization problem of (A1) and (A3)-(A4) since constraint (A2) would not be binding, which is equivalent to finding the intersection in the sequential *Step 3a*. Furthermore, the shadow price associated with (A3) will be the same with and without the non-binding constraint (A2), i.e., $\lambda_{SYS} = \lambda_C$. Therefore, the sequential clearing yields the same cleared quantities and clearing prices as the optimization problem (A1)-(A4).

- $\lambda_A\geq\lambda_{SYS}$. This means that product *A*'s cleared quantity in *Step 2* is greater than or equal to its quantity cleared in *Step 1*. As a result, the solution of *Step 1* optimization problem of (A1) and (A3)-(A4) violates constraint (A2). Therefore, constraint (A2) must be binding at the solution of the optimization problem (A1)-(A4), i.e., $Q_A = R_A$. Substitute this result into (A3), then the optimization problem (A1)-(A4) can be decomposed into the *Step 2* optimization problem for zone *A*, i.e., $\min_{Q_A} C_A(Q_A)$ subject to $Q_A \geq R_A$ with the shadow price associated with the constraint being λ_A; and a new optimization problem for zone *B*, i.e., $\min_{Q_B} C_B(Q_B)$ subject to $R_A + Q_B \geq R_C$ with the shadow price associated with the constraint being λ_B. The former optimization problem for zone *A* is equivalent to finding the intersection in *Step 2*, and the latter optimization problem for zone *B* is equivalent to finding the intersection in *Step 3b*. This proves the sequential clearing leads to the same cleared quantities as the optimization problem (A1)-(A4). By further comparing the KKT optimality conditions of the aforementioned two sequential optimization problems and the problem (A1)-(A4), it can be shown that $\lambda_A = \lambda'_A + \lambda_C$ and $\lambda_B = \lambda_C$. This proves the sequential clearing yields the same clearing prices as the optimization problem (A1)-(A4).

As a result, in both cases the sequential steps yield the same clearing result as solving (A1)-(A4). [*End of Proof*]

The above theorem proves the intuitive equivalence of the sequential clearing process and the cost minimization problem (A1)-(A4) with fixed requirements under the feasibility assumption.