XVIII ICIEOM 2012
International Conference on Industrial Engineering and Operations Management
Guimarães - Portugal
July 9 - 11, 2012

Using Clearing Function to Evaluate Pricing Capacity of Resources at Lows Levels of Utilization in Planning-Production

Raimundo J. B. De Sampaio*, Rafael R. G. Wollmann*

* Production Engineering Department, School of Engineering, Pontifical Catholic University of Parana, Campus of Prado Velho, Curitiba, Brazil

Email: raimundo.sampaio@pucpr.br; Rafael.wollmann@brturbo.com.br

Abstract

Theory suggests and daily industrial experience confirm that high levels of throughput in production depends on high levels of Work-In-Process (WIP) or releases into the system, and that sometimes increasing production capacity is of our best interest even before it becomes tight. However, the necessary information for this accomplishment is not provided by the most simple and used approach to this issue, which is the classical linear programming models. To circumvent this drawback in the approach of linear programming we develop here an approach based on the concept of clearing function, which strives with the approach of linear programming to allow pricing at low levels of utilization of capacity.

Keywords: Work-in-Process; Clearing Function; Utilization of Capacity; Dual Prices; Convex Programming.

1 Introduction

It is well known from the daily industrial experience that high levels of throughput in production depends on high levels of Work-In-Process (WIP) and releases into the system, and that high levels of WIP can increase the total lead time in systems, decreasing expected revenues (Hopp 2001, Asmundsson 2003). This clearly suggests that increasing production capacity might be of our best interest even before it becomes tight, as longer the performance of the system and not economy is the concern (Uzsoy 2011). However, the necessary information about when and how to do this accomplishment are not provided by the most used approach to this issue which is the classical Linear Programming (LP) models. In fact, as longer production planning is modeled using LP, WIP is not explicitly considered, but just a consequent balance among levels of production, final good inventory, and demand. Moreover, throughput accounts directly to nominal capacity and availability of resources ignoring others factors like lead time and workload of the system which contradicts both, the practical experience coming from flows shop, and the insights coming from queuing theory (Asmundsson 2009). The framework we develop here to modify the classical approach of LP is based on the concept of Clearing Function (CF) first suggest by Graves (Graves 86), Karmarkar (Karmarkar 89), and Srinivasan et alli (Srinivasan 1988), and more recently extensively used by Asmundsson et alli (Asmundsson 2009), Ali et alli (Ali 2010), Irdem et alli (Irdem 2010), and Mossbauer and Uzsoy (Uzsoy 2011), in the same setting considered here, just to name a few recent works. The approach we use to face this issue is twofold, in one side we assume that even though the behavior of CF function can be known, there is not in general an available analytical formulation to be used and only numerical information about medium number of throughput for any level of WIP are at hand. This assumption is mainly supported by the queuing theory which shows that in general it is not possible to explain analytically throughput as a function of WIP, and results from several different numerical experiments that shows a great deal of variability in throughput in response to the same level of WIP, which suggest that quite different analytic models could be used to represent throughput as a function of a measure of WIP, as longer it observes that the data distribution have a clearly standard concave behavior. However, in all cases the CF function could be approximate using a regression concave curve or a piecewise affine approximation accordingly its concave behavior, i.e., the angular parameters decrease from nominal production value (or from 1), to zero, and the linear parameter increase from zero (or from

a small value) to nominal production capacity. The second aspect is when we can manage to have an analytical formulation for CF function, which in particular, can come from a regression curve that adjusts the available data. In this case we price low levels of capacity throughout a convex programming model, and then we perform a sensitivity analysis for the model related to CF function using perturbation theory of convex duality to conclude about improving capacity. And this is a new achievement to the best of our knowledge. The remainder of this paper is organized as follow: In Section 2 a brief review of the literature is presented to shortly describe the related previous works on CF function and its rule in production-planning and pricing resources for improve capacity. In Section 3 the working model is built upon the classic LP model and the main assumptions on the model and on CF function as well, are discussed. In Section 4, some final conclusions are made.

2 Previous Related Works

Linear Programming (LP) is a widely used framework to address the problem of production-planning, which has been studied for several decades (Karmarkar 89, Graves 86, Hopp 2001, Jacob 2009, Ali 2010, Karl 2011). However, standard LP models presents a number of drawbacks in that their recommendations are inconsistent with the queuing behavior observed in most production facilities. One such drawback is that the dual variables associated with capacity constraints will only take nonzero values when the resource is fully utilized, contradicting both theoretical results from queuing theory and practical experience on the factory floor. Another serious drawback is related to lead time. Most LP models assume that a resource can maintain a constant lead time regardless of its workload, again contradicting basic queuing insights, actually implying infinite capacity. Practical experience and queuing theory show that the performance of production system is affected by the loading of the system related to its capacity, in particular the lead time, i.e., the mean time between the release of work for production and its completion, increase nonlinearly with increasing resources utilization (Hopp 2001, Graves 86, Karmarkar 89). Thus, deterministic models for production-planning suffer from a problem that in order to match demand and supply they need to consider lead times to plan releases, but in doing so they determine levels of resource utilization, which, in turn, determines the realized lead time of production. A number of iterative approaches have been proposed (Jacob 2009) to deal with this circularity. Those involve assuming initial fixed lead times to obtain a release plan, then simulating the release plan to obtain realized lead times estimates. These estimates are then used to generate a new release plan, until the procedure converges. A detailed review of the state of art in production-planning using piecewise linear approximation of CF function can be seen in (Uzsoy and Mossbauer 2011).

Clearing Function is an increasing bounded concave function that expresses the expected throughput of a capacitated resource over a period of time as a function of the average WIP level in the system over that period. Thus, the introduction of CF function into the production-planning model, while improving modeling the problem destroys its sweet linear structure which generally we likely want to preserve. The resulting convex problem, however, may be approximate to any degree of accuracy by a large scale LP problem which, in general, may requires decomposition technic to be solved due its scale size.

3 Production-Planning CF Model

The approach used here for modeling production using the role of CF function follows the paths of Ali (Ali 2010), Mossbauer and Uzsoy (Uzsoy 2011), Irdem et alli (Irdem 2010), which in its turns follow some previous works from Graves (Graves 86) and Karmarkar (Karmarkar 89). However, we use here a quite different assumption which is that we only have access to numerical information about the relationship between levels of WIP and throughput. The piecewise approach just mentioned intend to approximate the CF concave function by means of a bundle of piecewise affine functions, and so, model the problem as a modified LP problem, thus keeping the simplicity of LP models at the time that allows the possibility of pricing low levels of utilization of capacity using the standard LP sensitivity analysis. The approach used

here assumes that the analytical formulation for CF function comes from a regression curve which can be obtained from the available data about the behavior of throughput as a function of some measure of WIP. Now let's build the new CF model using the same problem addressed by Karmarkar (Karmarkar 89), for a single product produced in a single machine, defining the following set of parameters and decision variables: c_t is the unit cost of product at period t; h_t is the unit cost of handling inventory at period t; x_t is the level of production at the end of period t; I_t is the level of final good inventory at the end of period t; D_t is the demand to be satisfied by the end of period t; and C_t is the available nominal capacity at period t. The classic LP production-planning model is generally formulate as,

$$\begin{aligned} Minimize \quad & \sum_{t=1}^{T} (c_t\, x_t + h_t\, I_t) \\ s.t. \quad & I_t = I_{t-1} + x_t - D_t, \quad \forall t \\ & x_t \geq 0, I_t \geq 0, \quad \forall t. \end{aligned}$$

This model does not explicitly make any reference to WIP (W_t) and does not include releases (R_t) at the beginning of any period t as a decision variables, which turns the model to be of small practical value, since all the resources are supposed to be available in the beginning of first period, and lead time is constant, do not depending on the workload of the system. However, we may modify this LP model to incorporate WIP and releases, using the following direct extension,

$$\begin{aligned} Minimize \quad & \sum_{t=1}^{T} c_t\, x_t + h_t\, I_t + w_t W_t + r_t R_t \\ s.t. \quad & I_t = I_{t-1} + x_t - D_t \quad \forall t \\ & W_t = -x_t + W_{t-1} + R_t \quad \forall t \\ & x_t \geq 0, I_t \geq 0, W_t \geq 0, R_t \geq 0, \quad \forall t, \end{aligned}$$

which explicitly uses WIP and releases as decision variables. The new parameters r_t and w_t denote the unit cost of releases and WIP holding, respectively, and the new decision variables W_t and R_t denote respectively the amount of products in WIP and the releases at beginning of period t. To introduce the CF function, let $\delta(\widehat{W_t})$ be the concave CF function which governs production as a function of the workload of the system, where $\widehat{W_t} = W_{t-1} + R_t$, and links expected throughput x_t with levels of WIP and releases, thus implicitly bringing lead time to the scenario as a control variable of the workload of the system, instead of as a forecast variable, as usual. The CF model is a direct extension of classical model,

$$\begin{aligned} Minimize \quad & \sum_{t=1}^{T} c_t\, x_t + h_t\, I_t + w_t W_t + r_t R_t \\ s.t. \quad & I_t = I_{t-1} + x_t - D_t \quad \forall t \\ & W_t = -x_t + W_{t-1} + R_t \quad \forall t \\ & x_t - \delta(\widehat{W_t}) \leq 0 \quad \forall t \\ & x_t \geq 0, I_t \geq 0, W_t \geq 0, R_t \geq 0 \quad \forall t. \end{aligned}$$

And this is a convex programming model that, in general, we assume to satisfy the Slater's conditions.

3.1 Pricing low levels of capacity

The mathematical model defined above allows periods of congestion, non-congestion, and idle periods for the production plan, however we could impose initial conditions on $\boldsymbol{I_t}$, $\boldsymbol{W_t}$, and $\boldsymbol{R_t}$, in order for to drive the model for a specific state, specifically congestion periods. Now let go back to CF model which is a convex programming problem, and for sack of simplicity lets have our CF model generally written as,

$$\begin{aligned} Minimize \; & f_0(x) \\ s.t. \quad & f_i(x) \leq 0; \\ & h_j(x) = 0, \end{aligned}$$

where the constraints of non-negativity were incorporated into the inequality constraints, $\boldsymbol{f_0}(\boldsymbol{x})$ is a linear function, $\boldsymbol{f_i}(\boldsymbol{x})$ is a strictly convex function, and $\boldsymbol{h_j}(\boldsymbol{x})$ is an affine function, and assume that the constraints have a nonempty relative interior. The Lagrangian of this problem is,

$$l(x,\lambda,\mu) = f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \mu_j h_j(x), \quad \lambda_i \geq 0.$$

And its dual function is,

$$g(\lambda,\mu) = inf_x\, l(x,\lambda,\mu), \quad \lambda_i \geq 0,$$

which is always a non-smooth concave function since it is a piecewise infimum of a family of affine functions. If x is a feasible solution for the convex problem, then,

$$\forall(\lambda,\mu), \quad \lambda \geq 0, \quad g(\lambda,\mu) \leq f_0(x).$$

This important property, called weak duality, is easily verified since for all feasible solution for the convex problem we have that,

$$\sum_i \lambda_i f_i(x) + \sum_j \mu_j h_j(x) \leq 0, \qquad \forall\lambda_i \geq 0.$$

Furthermore, if x^* is an optimal solution for the convex problem and (λ^*,μ^*) an optimal solution for the so called dual problem,

$$\begin{array}{ll} Maximize & g(\lambda,\mu) \\ s.t. & \lambda \geq 0, \end{array}$$

then it follows that $g(\lambda^*,\mu^*) = f_0(x^*)$. The value $g(\lambda^*,\mu^*)$ gives the best lower bound for the convex problem, while the optimal value of primal problem, $f_0(x^*)$, gives the best upper bound for the dual problem. Since we intend to analyze the sensitivity of the optimal solution of primal problem (CF model) related to small variations of CF function let's consider small perturbations into the CF model in the following sense,

$$\begin{array}{lll} Minimize & f_0(x) & \\ s.t. & f_i(x) \leq u_i, & \forall i, \\ & h_j(x) = v_j, & \forall j, \end{array}$$

where u_i can be either positive, in which case the constrained $f_i(x)$ is relaxed, or negative, in which case the constraint $f_i(x)$ is tightened. According to CF model the perturbation u_i can be interpreted as variations over CF function, and since strong duality attains, this allow us to have valuables information on how the value of an optimal solution to CF model gives answer to perturbations of the CF function, which by its turns gives information on the value of increasing and decreasing capacity. Likewise, changes on the right side of equality constraints can be interpreted as variations over boundary conditions about demand, releases, and WIP, even though it doesn't matter to much here since we are mainly interested on how the CF model gives responses to small perturbations on the CF function to support decisions on increasing capacity. The Lagrangian function of the perturbed problem is,

$$L(x,\lambda,\mu) = f_0(x) + \sum_i \lambda_i (f_i(x) - u_i) + \sum_j \mu_j (h_j(x) - v_j), \qquad \lambda_i \geq 0,$$

and its dual function is,

$$G(\lambda,\mu) = inf_{(u,v)}\, L(x,\lambda,\mu), \quad \lambda_i \geq 0.$$

Let's define a set G of optimal values for the perturbed problem for each perturbation (u,v), and a function G^* over the set G, defined as $G^*(u,v) = inf_{(u,v)}\, G$. Observe that $G^*(0,0) = g(\lambda^*,\mu^*) = f_0(x^*)$ and from strong duality, for any pair $(x^*,(\lambda^*,\mu^*))$ of optimal solution,

$$g(\lambda^*, \mu^*) \leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_j \mu_j^* h_j(x^*)$$
$$\leq f_0(x^*) + \sum_i \lambda_i^* u_i + \sum_j \mu_j^* v_j$$
$$= f_0(x^*) + \lambda^{*T} u + \mu^{*T} v$$

which implies by concavity of $G^*(u,v)$ combined with the inclusion relation, $(-\lambda^*, -\mu^*) \in \partial G^*(0,0)$, that, $G^*(u,v) \geq G^*(0,0) - \lambda^T u - \mu^T v$. It should be realized that perturbation (u,v) account for variations of levels of utilization of capacity in the productive system, and so, perturbation theory can be viewed in this case as a sensitivity analysis about utilization of capacity. From the inequality $G^*(u,v) \geq G^*(0,0) - \lambda^T u - \mu^T v$, we can directly have several insight about the effect of perturbation on the CF function over the model, no matter whether the CF function is a strictly concave function or its piecewise linear approximation, since this insights comes from the general theory of convex duality, that is, comes from the quality of being convex duality and not from the kind of convex function used. Some of the most obvious insights are listed below as a matter of clarification.

1. If λ_i is large and the ith inequality is tight, which means that the perturbation $u_i < 0$, then $G^*(u,0)$ is guaranteed to greatly increased, i.e., the production costs greatly increase;
2. If λ_i is small and the ith inequality is tight, which means that the perturbation $u_i < 0$, then $G^*(u,0)$ may not increase too much, i.e., the production costs may or may not increase;
3. If λ_i is large and the perturbation $u_i > 0$ but small, then $G^*(u,0)$ is not guaranteed to decrease too much, i.e., the production costs is not guaranteed to decrease too much;
4. If λ_i is small and the perturbation $u_i > 0$ then $G^*(u,0)$ is not guaranteed to decrease too much, i.e., the production costs is not guaranteed to decrease too much.

We must bear in mind that $G^*(u,v)$ is the value of the dual function for the perturbed problem, which gives an inferior bound for the value of the primal function for each perturbation pair (u,v), and hence, if it increase, so does the primal function, which means at the end, that production costs increase. And by the contrary, if $G^*(u,v)$ decrease, since we have strong duality property, so decrease production costs. It is worthy to notice that this technology provides a sensitivity analysis for all possible cases, no matter if the CF function is a strictly concave function or just a concave function given by the infimum of piecewise affine functions.

4 Final Conclusions

As Karmarkar (Karmarkar 89) observed, out of some very simple cases it is not possible to have a closed expression for throughput in terms of WIP, even though it may exist. However the relationship between throughput and WIP may always be graphed for each product using queening model or simulation, always giving the same qualitative result, even though in the case of multiple products the output of any product can be affected by the WIP level of all others products as pointed out in Missbauer and Uzsoy (Uzsoy 2011), referring to a paper from Carey and Subrahmanian (Carey 2000). In general it is reasonable expect to access only numerical information about throughput as a function of WIP, and then a possible approach to this issue include to discretize the interval of throughput into several subintervals, obtain a regression line for each subinterval, and then construct a piecewise approximation for CF function to model the problem as a LP problem, or just obtain an overall regression concave function. In the case of piecewise linear approximation it can be chosen to varies the linear parameter which correspond to varies the piecewise linear approximation from inner to outer approximation, allowing sensitivity analysis. On the other side the problem can be treated directly as a convex optimization problem where the objective function is linear and the constraints define a convex non polyhedral set with nonempty relative interior whose CF function is obtained as a regression concave function overall intervals of WIP. Perturbations on the CF function provide the appropriated framework to sensitivity analysis, and feasible set with nonempty relative interior assure exact duality. Variations on the linear parameters in the case of piecewise linear approximation as well as perturbations on the CF function provides the same

environment for the general theory of convex perturbation to be applied, and perturbations account for levels of utilization of capacity.

References

(Ali 2010) Ali Kefeli; Reha Uzsoy; Yahya Fathi; Michael Kay, *Using a mathematical programming model to examine the marginal price of capacity resources*, To appear in Int. J. Production Economics.

(Carey 2000) Malachy Carey; Eswaran Subrahmanian, *An approach to modeling time-varying flows on congested networks*, Transportation Research, Part B, 34 (2000) 157-183.

(Srinivasan 1988) A. Srinivassan; M. Carey; T. E. Morton, *Resource Pricing and Aggregate Scheduling in Manufacturing Systems*, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA, 1988.

(Asmundsson 2009) Jacob M. Asmundsson; Ronald L. Rardin; Can Hulusi Turkseven; Reha Uzsoy, *Production Planning with Resources Subject to Congestion*, Published online 20 January 2009 in Wiley Interscience (www.interscience.wiley.com), DOI 10.1002/nav.20335.

(Asmundsson 2003) Jacob M. Asmundsson, **Tractable Nonlinear Capacity Models for Aggregate Production Planning**, Ph.D. Thesis submitted to the Faculty of Pardue University, May 2003.

(Rubin 90) David S. Ribin; Harvey M. Wagner, Shadow Prices: *Tip and Traps for Managers And Instructors*, Interface 20: July--August, 150-157, 1990.

(Irdem 2010) D. Fatih Irdem; N. Baris Kacar; Reha Uzsoy, Senior Member IEEE, *An Exploratory Analysis of Two Iterative Linear Programming-Simulation Approaches for Production Planning*, IEEE, Transactions on Semiconductor Manufacturing, Vol. 23, NO. 3, August 2010.

(Uzsoy 2011) Hubert Missbauer; Reha Uzsoy, *Optimization Models of Production Planning Problems*, **Planning Production and Inventories in the Extended Enterprise**, A State of Art Handbook, Volume 1, Edited by Karl G. Kempf, Pinar Keskinocak, Reha Uzsoy, Springer 2011.

(Graves 86) S. C. Graves, *A Tactical Planning Model for a Job Shop*, Operations Research, 34, 522-533, 1986.

(Salah 2011) Salah Elmaghraby, *Production Capacity: Its Bases, Functions and Measurements*, **Planning Production and Inventories in the Extended Enterprise**, A State of Art Handbook, Volume 1, Edited by Karl G. Kempf, Pinar Keskinocak, Reha Uzsoy, Springer 2011.

(Karmarkar 89) Uday S. Karmarkar, *Capacity loading and release planning with work-in-process and lead times,* Journal of Manufacturing and Operations Management, 2, 105-123, 1989.

(Hopp 2001) W. J. Hopp; M. L. Spearman, *Factory Physics: Foundations of Manufacturing Management*, Irwin/McGraw-Hill, Boston, 2001.