Stopping and allocation indices for piecewise deterministic flow processes

Max-Olivier Hongler
Département de Microtechnique, (DMT)
Institut de Production Microtechnique, (IPM)
E.P.F.L.
CH-1015 LAUSANNE.

Abstract

We consider the optimal stopping problem for a class of continuous time random evolutions described by stochastic differential equations with alternating renewal processes as noise sources. Such random processes are naturally encountered when the fluid limit of production flows is considered. The explicit solution of the stopping problem enables to derive Gittins indices in closed forms and then to discuss the associated dynamic allocation problem.

Keywords: Dynamic allocation of jobs, optimal stopping, piecewise-deterministic processes, continuous time Gittins indices.

1 Introduction

The scheduling of multiclass make-to-stock facility with a limited resource, operating in a fluctuating environment, is a favorite subject in the ongoing research devoted to manufacturing systems, [Wei 92], [Vea 96], [Ha 97], [Yao 97], [Pen 97], [Fed 98], [Fed 99], [Ver 99]. Due to the problem complexity it is, in general, impossible to analytically derive exact optimal production policies. Moreover, even when numerical results are available, the optimal solutions are often too complex to be implemented in actual situations. Therefore, one heavily relies on approximation schemes yielding efficient though, in general, only sub-optimal policies. When set-up costs and/or times required to change from one production class to another are negligible, efficient approximation schemes can be obtained by using priority indices $M_j(X_j(t))$, $j = 1, 2, \ldots, d$, with d denoting the numbers of items present. These indices $M_j(X_j(t))$ indicate the relative urgency to produce the particular item j when its time-dependent surplus level is at position $X_j(t)$. Using such priority indices, the resulting scheduling policy reads for example as: *at each time produce the item having the lowest* $M_j(X_j(t))$.

For a rather large class of systems, scheduling policies based on priority indices may actually be optimal. *For which class of dynamics is this realized ?* This fundamental question has been addressed and answered in a recent series of papers by [Sha 92], [Ber 96] and [Yao 97]. The essential property necessary for index policies to hold is the presence of conservation laws associated with the underlying dynamics. As an illustration, the scheduling problems which can be modeled by multi-armed bandit processes are indexable and belong to above mentioned class of dynamics [Ber 96]. Hence once the explicit knowledge of the priority indices $M_j(X_j(t))$ is available, the optimal scheduling can be explicitly derived for this class of dynamics. For random evolutions in continuous time, several recent contributions have been devoted to the calculation of the indices $M_j(X_j(t))$. These papers were stimulated by the pioneering and elegant paper of I. Karatzas [Kar 84] where explicit analytic expressions were derived for the indices associated with diffusion random processes. The aim of the present note is to calculate the priority indices for a simple, non-diffusive class of dynamics. In this class of models the noise source is a markovian alternating renewal process which occur in the description of failure prone machines, [Bie 88].

2 Optimal stopping for a simple, non-diffusive, stochastic process

To calculate the priority indices, we follow the method introduced by [Whi 82] and used by [Kar 84]. Accordingly, we first introduce and solve an associated optimal stopping problem defined for a class of scalar stochastic processes, (i.e. a single product in the manufacturing context). The class random evolutions studied below is commonly use in the fluid limit modeling of manufacturing flows delivered by failure prone machines, (see for instance [Bie 88] and [Ger 94].)

2.1 A simple class of dynamics

Let us consider the scalar time evolution for $t \geq 0$ given by:

$$\frac{d}{dt}X(t) = I(t), \;\; X(0) = x, \;\; I(0) = i, \;\; \text{and} \;\; X(t) \in R, \tag{1}$$

where $I(t)$ is an alternating renewal process taking the alternate constant values U, (up) and D, (down). Assume that the sojourn times in the U and D phases are exponentially distributed with parameters λ and μ respectively. We shall focus our attention to the cases for which we have:

$$D\lambda + U\mu > 0 \;\; \text{with} \;\; D < 0 \;\; \text{and} \;\; D + U > 0. \tag{2}$$

so that the asymptotic average of $X(t)$ increases with time. The class of picewise deterministic evolutions of the type given by Eq.(1) is extensively discussed in [Pin 91].

2.2 The associated optimal stopping problem

Now we introduce a simple optimal stopping problem defined by the cumulate reward function $J(x, i)$:

$$J_{(x,i)} = E_{(x,i)} \int_0^{\tau} h\left(X(s)\right) e^{-\alpha s} ds + m e^{-\alpha \tau}, \quad m \in R^+. \tag{3}$$

where τ is a stopping time, m a terminal reward and α the discount a discount factor. The instant reward function $h(x)$ is a twice differentiable function, it is strictly increasing. Additionally, it has the following asymptotic properties:

$$\lim_{x \to \infty} h(x) = \alpha K, \quad \lim_{x \to -\infty} h(x) = \alpha k, \quad k < K, \tag{4}$$

with $\alpha > 0$ and:

$$\lim_{|x| \to \infty} \frac{d}{dx} h(x) = 0. \tag{5}$$

In Eq.(5),the notation $E_{(x,i)}$ stands for the mathematical expectation when the initial conditions are $X(t = 0) = x$ and $I(t = 0) = i$. The problem is to find an optimal stopping time τ^* which maximizes the expected reward $J_{(x,i)}$.

To solve this problem, we note that when the process $(X(t), I(t))$ given in Eq.(1) is in a given state, we can either wait a certain period of time before stopping or stop immediately. The resulting rewards are calculated by assuming that the subsequent decisions are optimal. Accordingly, we can write:

- i) If at time t, the decision is to wait for an (infinitesimal) time ξ before stopping, the optimal reward $J_{(X(t),I(t))}$ is greater than the running reward from t to $t+\xi$ and the optimal reward from time $t+\xi$ onwards. This yields the inequality:

$$J_{(X(t),I(t))} \geq \int_t^{t+\xi} e^{-\alpha(s-t)} h\left(X(s)\right) ds + e^{-\alpha\xi} J_{(X(t+\xi),I(t+\xi))}. \tag{6}$$

- ii) If the decision at time t is to stop immediately, the reward is precisely: $J_{(X(t),I(t))} = m$

Let us focus on case i) for which we consider two possible rewards depending on the realizations of the process $I(t)$, namely: $I = U$ or $I = D$. To this two possibilities, we associate the rewards functions $J_{(x,U)}$ and $J_{(x,D)}$ respectively. With this notation and the markov character of the alternating process $I(t)$, the first order time expansion of Eq.(6) yields:

$$0 \geq h(x) - (\alpha+\lambda) J_{(x,U)} + U \frac{d}{dx} J_{(x,U)} + \lambda J_{(x,D)} \tag{7}$$

and

$$0 \geq h(x) - (\alpha+\mu) J_{(x,D)} + D \frac{d}{dx} J_{(x,D)} + \mu J_{(x,U)}. \tag{8}$$

The strictly increasing nature of $h(x)$ suggests to distinguish between three regimes holding in three adjacent intervals $\mathcal{O}_j$, $j = 1, 2, 3$ defined by two thresholds $b_u < b_d$. Accordingly, we shall write:

- a) $\mathcal{O}_1 = [b_d, \infty)$
- b) $\mathcal{O}_2 = [b_u, b_d]$
- c) $\mathcal{O}_3 = (-\infty, b_u]$

Now we construct the reward functions $J^{(k)}_{(x,D)}$ and $J^{(k)}_{(x,U)}$ for $k = 1, 2, 3$ holding in the intervals $\mathcal{O}_1$, $\mathcal{O}_2$, and $\mathcal{O}_3$ respectively. The qualitative behavior of the reward functions is sketched in Figure 1.

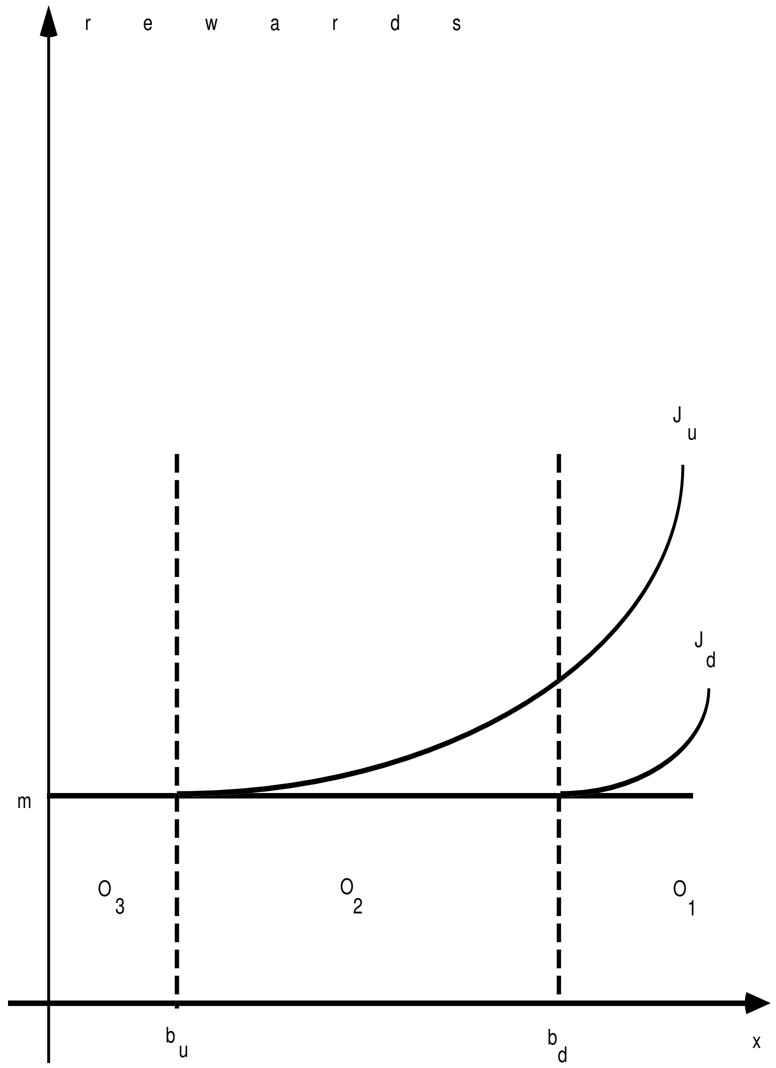

Figure 1: Qualitative behavior of the rewards functions $J_{(x,U)}$ and $J_{(x,D)}$ in the intervals $\mathcal{O}_k,\ k = 1, 2, 3$

- a) *Solution in the interval* $\mathcal{O}_1$. This is a continuation region for both $J^{(1)}_{(x,D)}$ and $J^{(1)}_{(x,U)}$. In this region Eqs.(7) and (8) hold simultaneously with equality. We can therefore combine Eqs.(7) and (8) to get two second order differential equations having the form:
$$\mathcal{A}\left(J^{(1)}_{(x,U)}\right) + h_D(x) = 0 \tag{9}$$
and
$$\mathcal{A}\left(J^{(1)}_{(x,D)}\right) + h_U(x) = 0, \tag{10}$$
where $\mathcal{A}$ stands for the differential operator:
$$\mathcal{A}(\cdot) = -UD\frac{d^2}{dx^2}(\cdot) + \delta\frac{d}{dx}(\cdot) - \alpha\rho(\cdot) \tag{11}$$
with
$$\delta = D(\alpha+\lambda) + U(\alpha+\mu) > 0, \quad \text{and} \quad \rho = (\alpha+\mu+\lambda) \tag{12}$$
and
$$h_D(x) = \rho h(x) - Dh'(x) \quad \text{and} \quad h_U(x) = \rho h(x) - Uh'(x), \tag{13}$$
with the notation $\frac{d}{dx}h(x) = h'(x)$.
For later use, we shall restrict our attention to the class $\mathcal{H}$ of strictly increasing functions $h_U(x)$, namely:
$$h_U(x) \in \mathcal{H} \iff h'_U(x) = \rho h'(x) - Uh''(x) > 0. \tag{14}$$
The solutions of the linear differential Eqs.(9) and (10) are easy to derive. Identical equations arise in [Kar 84] and using these results, we directly obtain:
$$J^{(1)}_{(x,D)} = A_{d,1}\exp\{-\gamma_+ x\} + p_d(x), \quad \text{for} \quad x > b_d \tag{15}$$
and
$$J^{(1)}_{(x,U)} = A_{u,1}\exp\{-\gamma_+ x\} + p_u(x), \quad \text{for} \quad x > b_d \tag{16}$$
with $A_{d,1}$ and $A_{u,1}$ being integration constants and:
$$\gamma_+ = \frac{-1}{2DU}\left[\sqrt{\delta^2 - 4DU\rho\alpha} + \delta\right] > 0, \tag{17}$$

The functions $p_d(x)$ and $p_u(x)$ are particular solutions of Eqs.(9) and (10). They represent the rewards when running forever, (i.e. when no stopping occurs). Hence, we have $p_u(x) > k$ and $p_d(x) > k$. In our case, two particular solutions of the inhomogeneous Eqs.(9) and (10), respectively:

$$p_u(x) = E_{(x,U)} \int_0^\infty e^{-\alpha t} h\left(X(t)\right) dt$$

$$= \mathcal{N}\left[e^{-\gamma_+ x} \int_{-\infty}^{x} h_D(y) e^{\gamma_+ y} dy + e^{\gamma_- x} \int_x^\infty h_D(y) e^{-\gamma_- y} dy\right] \tag{18}$$

and

$$p_d(x) = E_{(x,D)} \int_0^\infty e^{-\alpha t} h\left(X(t)\right) dt$$

$$= \mathcal{N}\left[e^{-\gamma_+ x} \int_{-\infty}^{x} h_U(y) e^{\gamma_+ y} dy + e^{\gamma_- x} \int_x^\infty h_U(y) e^{-\gamma_- y} dy\right], \tag{19}$$

with:

$$\mathcal{N} = \left[-DU(\gamma_+ + \gamma_-)\right]^{-1} \tag{20}$$

and

$$\gamma_- = \gamma_+ + \frac{\delta}{UD} \leq 0. \tag{21}$$

- b) *Solution in the interval $\mathcal{O}_2$.* This is a stopping region when $I(t) = D$ and hence $J_{(x,D)} = m$ and a continuation region when $I(t) = U$. Hence the use of Eq.(7) with $J_{(x,D)} = m$ immediately yields:

$$J^{(2)}_{(x,U)} = e^{\omega x}\left[A_{u,2} - \int_0^x \left[\frac{h(\xi) + \lambda m}{U}\right] e^{-\omega \xi} d\xi\right], \tag{22}$$

  with $A_{u,2}$ being an integration constant and $\omega = \frac{\alpha+\lambda}{U}$.

- c) *Solution in the interval $\mathcal{O}_3$.* This is a global stopping region and we therefore have $J^{(3)}_{(x,U)} = J^{(3)}_{(x,D)} = m$

From Eqs.(15), (16), (21), we see that we have to determine three constants of integration $A_{u,1}$, $A_{u,2}$ and $A_{d,1}$. Moreover the thresholds b_u and b_d are yet unknown. Hence, we need five relations to determine these five unknowns. These relations are:

- i) $J^{(3)}_{(b_d,D)} = m$.
- ii) $\frac{d}{dx} J^{(3)}_{(x,D)} \mid_{(x=b_d)} = 0$, smooth fitting in b_d.
- iii) $J^{(2)}_{(b_u,U)} = m$.
- iv) $\frac{d}{dx} J^{(2)}_{(x,U)} \mid_{(x=b_u)} = 0$, smooth fitting in b_u.
- v) $J^{(1)}_{(b_d,U)} = J^{(2)}_{(b_d,U)}$.

The relations i) and ii), yields calculations identical with those exposed in [Kar 84], (see in particular Eqs.(3.14) and (3.20) in [Kar 84]). Accordingly, we obtain:

$$\begin{cases} M(b_d) = m \\ A_{d,1} = \frac{p'_d(b_d)}{\gamma_+} \exp(\gamma_+ \, b_d) \end{cases} \tag{23}$$

with:

$$M(x) = \frac{1}{\alpha} \int_0^\infty h_U(x + \frac{z}{\gamma_-}) e^{-z} dz \tag{24}$$

The relations iii) and iv) imply:

$$\begin{cases} h(b_u) = \alpha \, m \\ A_{u,2} = m \, e^{-\omega b_u} + \int_0^{b_u} \left[\frac{h(\xi) + \lambda m}{U}\right] e^{-\omega \xi} d\xi. \end{cases} \tag{25}$$

Finally, the relation v) gives.

$$A_{u,1} = e^{\gamma^+ b_u} \left[m e^{\omega(b_d - b_u)} - \int_{b_u}^{b_d} \left[\frac{h(\xi) + \lambda m}{U}\right] e^{-\omega \xi} d\xi - p_u(b_d) \right] \tag{26}$$

As we imposed in Eq.(14) $h_U(x)$ to be strictly increasing, the function $M(x)$ defined in Eq.(23) is itself strictly increasing, (see Eq. (3.19) in [Kar 84] for the direct demonstration). Hence, the parameters b_d and b_u are the unique solutions of Eqs.(22) and (24) respectively.

In addition to the property given in Eq.(3), note that the following properties hold:

$$\lim_{x \to \infty} M(x) = K \tag{27}$$

and similarly:

$$\lim_{x \to -\infty} M(x) = k. \tag{28}$$

It is important to realize that the functions $M(x)$ and $h(x)$ represents the smallest values of the terminal reward m which makes immediate

stopping profitable if the process is in the states (x, D) and (x, U) respectively. This is precisely the definition of the dynamic allocation index introduced by Gittins, [Git 89].
In view of the above results, we can now formulate:

Theorem. Assume that the instant reward function $h_U(x)$ is strictly increasing. Then for the initial conditions $X(t=0) = x$ and $I(0) = D$ respectively $I(0) = U$, the optimal stopping times are $\tau_d^*(x, m)$ and $\tau_u^*(x, m)$ read:

- a) $k < m < K$.

$$\tau_d^*(x, m) = \inf \{t \geq 0; X(t) \leq b_d\} \tag{29}$$

  and

$$\tau_u^*(x, m) = \inf \{t \geq 0; X(t) \leq b_u\} \tag{30}$$

  where the thresholds b_d and b_u are implicitly given in Eqs.(22) and (24).

- *b*) $m > K$.
  In this case we have: $b_d = b_u = \infty$ and therefore $\tau_d^*(x, m) = \tau_u^*(x, m) = 0$.
  In this case, the associated reward functions read $J^{(3)}(x, U) \equiv J^{(3)}(x, D) = m$

- *c*) $m < k$. In this cases we have: $b_d = b_u = -\infty$ and hence $\tau_d^*(x, m) = \tau_u^*(x, m) = \infty$.
  In this case, the associated reward functions read:

$$J^{(1)}(x, D) = p_d(x) > k \quad \text{and} \quad J^{(1)}(x, U) = p_u(x) > k.$$

Remark: When $k < m < K$, Eqs.(28) and (29) implies that at the stopping time $\tau_d^*(x, m)$ or $\tau_u^*(x, m)$, the process $I(t)$ will always be in the D phase, namely:

$$I(\tau_d^*(x, m)) = D \quad \text{or} \quad I(\tau_u^*(x, m)) = D \tag{31}$$

2.3 Dynamic allocation problem

Let us now consider the dynamic system characterized by:

$$\frac{d}{dt} X_j(t) = \mathbf{1}_{[\chi(t)=j]} I_j(t), \quad \chi, j = 1, 2, ..., d. \tag{32}$$

where $\mathbf{1}_{[\chi(t)=j]}$ is the indicator function, (i.e. $\chi(t) \in \{1, 2, \ldots, d\}$) and $I_j(t)$, $j \in \{1, 2, \ldots, d\}$ are independent alternating renewal processes similar to the one introduced in Eq.(1) with states U_j and D_j, $j \in \{1, 2, \ldots, d\}$. In other words, the process $\chi(t)$ indicates which project is engaged at time t. Remember that when a project is not engaged it remains "frozen". The initial conditions for Eq.(31) read:

$$\vec{x} = (X_1(t=0), X_2(t=0), ..., X_d(t=0))$$

and

$$\vec{i} = (I_1(t=0), I_2(t=0), ..., I_d(t=0)) \,. \tag{33}$$

The problem is to find an optimal allocation policy $\mathcal{P}^* = \{\chi^*(t)\}$ which maximizes the expected reward given by:

$$J(\vec{x}, \vec{i}; \mathcal{P}^*) = E_{\vec{x},\vec{i}} \int_0^\infty e^{-\alpha t} h\left(\chi^*(t), X_{\chi^*(t)}(t)\right) dt. \tag{34}$$

The reward functions $h(i, x)$, for $i = 1, 2, ..., d$, are assumed to belong the the class $\mathcal{H}$, defined in Eq.(14), for $j = 1, 2, \ldots, d$. If $\chi(t) = j$, one acquires an instant reward $h(j, X_j(t))$ per unit time, discounted by a factor $e^{-\alpha t}$. Moreover as in Eq.(3), we impose:

$$\lim_{x\to\infty} h(j, x) = \alpha K, \quad \lim_{x\to-\infty} h(j, x) = \alpha k, \;\; \forall\, j = 1, 2, \ldots d\,; \;\; k < K, \tag{35}$$

Using the fact $M_j(x)$ and $h(j, x)$ for $j = 1, 2, \ldots, d$ are dynamic allocation indices, it follows directly that the optimal allocation policy $\mathcal{P}^*$ will be given by a "play the leader policy" of the form continuing the project with the leading index, we namely have (see also [Kar 84]):

$$\chi^*(t) = \min\left\{\{1 \le q \le d\}, \mathcal{M}_q\left(X_q(t)\right)\right\}, \tag{36}$$

with

$$\mathcal{M}_q\left(X_q(t)\right) = \begin{cases} \max_{\{1\le j\le d\}} M_j\left(X_j(t)\right) & \text{if } I_j(t) = D; \\ \max_{\{1\le j\le d\}} h\left(j, X_j(t)\right) & \text{if } I_j(t) = U. \end{cases} \tag{37}$$

3 Conclusion and perspectives

Quite few are the random evolution processes for which an explicit solution of the optimal stopping problems of the type Eq.(5) are possible. Such class of models are worth to be worked out. This is even more so when the underlying dynamics is intimately related to actual applications. The stochastic differential equation of the type given by Eq.(1) arises naturally in the fluid modeling of the production flows delivered by failure prone machines. Hence, using the Gittins indices, the model discussed here offer several possibilities to explicitly discuss the dynamic allocation problem for flexible production flows in the fluid modeling context.

Among several possibilities to extent the study proposed in this short and prospective note, we would like to mention the possibility to relax the markovian character of alternating renewal process which models the noise source in (1). This can be achieved by introducing the method of the supplementary variable commonly used in the context of queueing theory

4 References

[Ber 96] D. Bertsimas and J. Niño- Mora. Conservation laws, extended polymatroids and multi-armed bandit problems ; A polyhedral approach to indexable systems.

[Bie 88] T. Bielicki and T. R. Kumar. "Optimality of zero inventory policies for unreliable manufacturing systems" Oper. Res bf 36, (1988) 532-541.

[Fed 99] A. Federgruen and Z. Katalan. The impact of adding a make-to-order item to a make-to-stock production system. Management Sci. **45**, 980-994, (1999)

[Ger 94] S. B. Gershwin *Manufacturing Systems Engineering.* Prentice Hall (1994).

[Git 89] J. C. Gittins. *Multi-Armed Bandits Allocation Indices.* J. Wiley (1989).

[Ha 97] A. Ha. Optimal dynamic scheduling policy for a make-to-stock production system. Operation Res.**45**, 42-54, (1997)

[Kar 84]. I. Karatzas. "Gittins indices in the dynamic allocation problem for diffusion processes". Ann. of Probab. **12**, (1984), 173-192.

[Pen 97] A. Peña and P. Zipkin. Dynamic scheduling rules for a multiproduct make-to-stock queue. Operation Res. **45**, 919-930, (1996).

[Pin 91]. M. A Pinsky. *Lectures on Random Evolution.* World Scientific, (1991).

[Sha 92] J. G. Shanthikumar and D. D. Yao. Multiclass queuing systems: Polymatroid structures and optimal scheduling control. Oper. Res. **40** , (1992), S293-S299.

[Vea 96] M. H Veatch and L.M. Wein. Scheduling a make to stock queue: Index policies and hedging points. Operation Res. **44**, 634-647, (1996).

[Ver 99] F. de Véricourt, F. Karaesmen and Y. Dallery. Dynamic scheduling in a make-to-stock system: A partial Characterization of optimal policies. To appear in Operation Res.

[Wei 1992] P. Wien. Dynamic scheduling of a multi class make-to-stock queue. Operation Res.**40**, 724-735, (1992)

[Whi 82] P. Whittle "*Optimization over Time. Dynamic Programming and Stochastic Control*" J. Wiley and Sons, New-York (1982).

[Yao 97] D. D. Yao and Li Zhang. Stochastic Scheduling via polymatroid optimization. In Lecture in Applied Mathematics (Eds. G.G. Yin and Q. Zhang), Vol 33, (1997), 333-364.