STABILITY AND SINGULAR PERTURBATIONS IN CONSTRAINED MARKOV DECISION PROBLEMS

Eitan Altman
INRIA
Centre Sophia Antipolis
06565 Valbonne Cedex, France

Vladimir A. Gaitsgory
Department of Economics
and Business Administration
Bar-Ilan Univ., Ramat-Gan, Israel

Abstract

We consider CMDPs (Constrained Markov Decision Problems) with the average cost criterion and a single ergodic chain, or the discounted cost with a general multi-chain structure. We establish new conditions for stability of the optimal value and control to changes of the parameters of the problem, such as immediate costs, transition probabilities and the discount factor. We then study singular constrained problems, for which the optimal value and controls exhibit discontinuities.

1 Introduction

In recent years growing attention has been given to CMDPs [3, 8]. Such problems frequently arise in telecommunications [9]. The question of sensitivity of CMDPs arises in the modeling of systems, where the values of some parameters are not precisely known, or in adaptive control problems [1]; there one typically uses on-line estimates of unknown parameters and updates the control accordingly. One is interested to know whether the convergence of estimates to the true parameters implies the convergence of controls and optimal value. As the solution of CMDPs is obtained by Linear Programming (LP) methods, the theory of sensitivity of LPs [4] was used in [1, 2] to establish conditions for continuity properties in different parameters under (1) "Slater type" condition, i.e. the existence of a policy for which all the constraints are achieved with strict inequality, and (2) assumption **A1**: under any stationary deterministic policy, the state space consists of a single ergodic class.

In some cases the Slater condition is too restrictive, e.g. when there are also equality constraints. In Section 4 we obtain simple alternative sufficient conditions for stability of CMDPs. In Section 5 we investigate singular perturbations of the discount factor and transition probabilities. For motivation for studying such problems in non constrained MDPs see [5, 6]. The source of singularities in the problems introduced there is the ergodic structure; they consider the case of N disjoint recurrent classes, with a perturbation in the transition matrix that allows the passage with low probability between the classes. In our case the source of singularities is the extra constraints in the MDP. We use the theory of singular perturbations of LPs [7] to compute the limit of the optimal value and policy as the perturbation vanishes.

2 The Model

Let $\{X_t\}_{t=0}^{\infty}$ be the discrete time state process, defined on the *state space* $\mathbf{X} = \{1, ..., N\}$; the action A_t at time t takes values in the finite *action space* $\mathbf{A}$. Denote the *history* process by $\mathcal{H}_t := (X_0, A_0, ..., X_t, A_t)$. The dynamics is given by $P_{xay} := P(X_{t+1} = y \mid X_t = x; A_t = a) = P(X_{t+1} = y \mid \mathcal{H}_{t-1} = h, X_t = x; A_t = a)$. A policy u in the *policy space* U is a sequence $u = \{u_0, u_1, ...\}$, where $u_t(\cdot \mid \mathcal{H}_{t-1}, X_t)$, applied at time t, is a conditional probability over $\mathbf{A}$. Denote the probability measure corresponding to u and initial distribution γ on $\mathbf{X}$ by P_γ^u. (When γ is concentrated on a single state x we use x instead of γ in the above notation). Let $U(S)$ and $U(D)$ denote the *stationary (randomized)* and *stationary deterministic policies* respectively. For $g \in U(S)$, let $p_{a|x}^g := u_t(A_t = a \mid X_t = x)$. Under g, X_t is a Markov chain with transition probabilities $P_{xy}^g := \sum_{a \in \mathbf{A}} p_{a|x}^g P_{xay}$. Define $\hat{U}_\gamma$ the set of policies for which a single limit exists for $t^{-1} \sum_{s=0}^{t} P_\gamma^u(X_s = y, A_s = a)$ as $t \to \infty$, for each y, a.

Denote $\delta_a(x)$ the Kronecker delta. For any finite set B, denote by $|B|$ the number of elements in B. For $D, V \in \mathbb{R}^K$, the notation $D < V$ stands for $D_k < V_k$, $k = 1, \ldots, K$, with a similar convention for matrices. For two matrices ζ, Q of appropriate dimensions, $\zeta \cdot Q$ stands for summation over common indices (scalar product). Q^T denotes the transposed of Q.

Let $C(\gamma, u)$ and $D(\gamma, u) := \{D^k(\gamma, u)\ ,\ 1 \le k \le K\}$ (defined below) be cost functions associated with policy u and initial distribution γ on $\mathbf{X}$. Call u *feasible* if for a given $V \in \mathbb{R}^K$, $D(\gamma, u) \le V$. Define (**COP**): Find a feasible $u \in U$ that minimizes $C(\gamma, u)$.

Let $c(x, a)$, $d(x, a) := \{d^k(x, a)\ ,\ k = 1, ..., K\}$ be real valued instantaneous cost functions. Let $0 < \beta \le 1$ be a discount factor. Introduce the *normalized discounted* cost functionals:

$$\begin{aligned} C_\beta(\gamma, u) &:= \overline{\lim}_{t\to\infty} (\textstyle\sum_{s=0}^{t} \beta^s)^{-1} E_\gamma^u \left[\sum_{s=0}^{t} \beta^s c(X_s, A_s) \right] \\ D_\beta^k(\gamma, u) &:= \overline{\lim}_{t\to\infty} (\textstyle\sum_{s=0}^{t} \beta^s)^{-1} E_\gamma^u \left[\sum_{s=0}^{t} \beta^s d^k(X_s, A_s) \right] \qquad k = 1, ..., K \end{aligned} \tag{1}$$

When $\beta = 1$, (1) reduces to the standard expected average cost. For $\beta < 1$, (1) can be written as the *standard discounted* costs with a normalization constant: $C_\beta(\gamma, u) = (1 - \beta) E_\gamma^u \left[\sum_{s=0}^{\infty} \beta^s c(X_s, A_s) \right]$ where $D_\beta^k(\gamma, u)$ is defined similarly. COP_β is defined by using the the costs (1) respectively in the general definition of COP.

For the expected average case define for $g \in U(S)$ the vectors of relative cost: $H^0(x,g) \stackrel{\text{def}}{=} \sum_{s=0}^{\infty} E_x^g \left[c(X_s,A_s) - C_1(x,g)\right]$ and $H^k(x,g) \stackrel{\text{def}}{=} \sum_{s=0}^{\infty} E_x^g \left[d^k(X_s,A_s) - D_1^k(x,g)\right]$, $k = 1, ..., K$.

3 Computation of optimal policies for CMDPs

3.1 CMDPs and Linear Programs

Consider $\mathbf{LP}_\beta$: Find $z^* := \{z^*(y,a)\}_{y,a}$ that minimizes $\mathcal{C}(z) := c \cdot z$ s.t.

$$\sum_{y,a} z(y,a)\left[\delta_v(y) - \beta P_{yav}\right] = [1-\beta]\gamma(v) \qquad v \in \mathbf{X} \tag{2}$$

$$\mathcal{D}^k(z) := d^k \cdot z \le V_k \qquad 1 \le k \le K \tag{3}$$

$$\sum_{y,a} z(y,a) = 1 \tag{4}$$

$$z \ge 0 \tag{5}$$

Given nonnegative numbers $z = \{z(y,a)\}_{y,a}$ define $g(z) \in U(S)$ by $p_{a|y} = z(y,a)[\sum_a z(y,a)]^{-1}$. If the denominator is zero then $g(z)$ is defined arbitrarily.

Lemma 1 *([8]) Assume $\beta < 1$ or A1. Then (i) COP is feasible iff LP_β is; if COP_β is feasible then there exists an optimal stationary policy.*
(ii) For every $u \in U(S)$, $\zeta(y,a) \stackrel{\text{def}}{=} \lim_{t\to\infty}[\sum_{s=0}^t \beta^s]^{-1} \sum_{s=0}^t \beta^s P_\gamma^u(X_s = y) p_{a|y}^u$ satisfies (2) (4) and (5). (iii) If $\beta < 1$ or $u \in U(S)$, then $\mathcal{C}(\zeta) = C_\beta(\gamma,u)$ and $\mathcal{D}(\zeta) = D_\beta(\gamma,u)$. If u is feasible for COP_β then ζ satisfies (3). (iv) Pick any optimal solution z^* of LP_β. Then the optimal value for COP is equal to $\mathcal{C}(z^*)$; $g(z^*)$ is optimal for COP.

Remark: Consider COP with inequality constraints replaced by equality constraints. Then Lemma 1 still holds if $\beta < 1$, but (i) and (iv) do not hold generally for $\beta = 1$ even under A1. (Of course the appropriate inequality constraints (3) in LP_β should also be replaced by equality constraints.) Moreover, for the latter case COP may be feasible and yet there may not exist a feasible stationary policy (see Counterexample 2.7 in [3] for a related example). These problems arise in the expected average case due to the limsup in the definition of the costs, from which it follows that an equality constraint cannot be expressed as two inequality constraints. However, Lemma 1 holds for $\beta = 1$ if COP is restricted to $\hat{U}_\gamma$.

Lemma 2 *Assume $\beta < 1$. Then the constraints (2) are linearly independent.*

Proof: Assume (i) $\gamma(y) > 0$ for all $y \in \mathbf{X}$ (ii) there are $L < |\mathbf{X}|$ linearly independent equations among (2). Then there exists z satisfying (2), (4), (5) with at most $|\mathbf{X}| - 1$ non zero components. However, this contradicts the fact that $\gamma(y) > 0$ since $\sum_a z(y,a) \geq (1-\beta)\gamma(y)$ (see (2)). ∎

Remark: Assume A1 and $\beta = 1$. Then [2] there are $|\mathbf{X}| - 1$ linearly independent equations among (2). For $\beta < 1$, (4) is redundant and obtained by summing (2) over all $v \in \mathbf{X}$.

We shall use the following generic notation for LP_β. Define the matrix A of size $\{|\mathbf{X}| + K + 1\} \times \{|\mathbf{X}| \times |\mathbf{A}|\}$ as $A = \left(A^{(1)}, A^{(2)}, A^{(3)}\right)^T$ where $A^{(1)}$ is of size $\{|\mathbf{X}|\} \times \{|\mathbf{X}| \times |\mathbf{A}|\}$ with $A^{(1)}(v,(y,a)) \stackrel{\text{def}}{=} \delta_y(v) - \beta P_{yav}$; $A^{(2)}$ is of size $\{K\} \times \{|\mathbf{X}| \times |\mathbf{A}|\}$ with $A^{(2)}(k,(y,a)) \stackrel{\text{def}}{=} d^k(y,a)$; and $A^{(3)}$ is of size $\{1\} \times \{|\mathbf{X}| \times |\mathbf{A}|\}$, with $A^{(3)}(v,(y,a)) \stackrel{\text{def}}{=} 1$. Define the vector $b = \left(b^{(1)}, b^{(2)}, b^{(3)}\right)^T$ of size $\{|\mathbf{X}| + K + 1\}$, where $b^{(1)}$, of size $\{|\mathbf{X}|\}$, is given by $b^{(1)}_v = (1-\beta)\gamma(v)$; $b^{(2)}$, of size $\{K\}$, is given by $b^{(2)}_k = V_k$ and $b^{(3)} = 1$ is of size $\{1\}$. LP_β is thus written as: minimize $c \cdot z$ s.t. $A^{(1)}z = b^{(1)}$, $A^{(2)}z \leq b^{(2)}$, $A^{(3)}z = b^{(3)}$, $z \geq 0$.

3.2 Conditions for stability of LPs

We summarize below conditions for the stability of LPs. Let $H_n \in \mathbb{R}^l$. $\underline{\lim}_n H_n$, $\overline{\lim}_n H_n$ and $\lim_n H_n$ are defined as in [4]. Given any function σ from $\mathbb{R}^l$ to $\mathbb{R}$ and a set $H \subset R^l$, define $M(\sigma|H)$ the subset of H where σ achieves its minimum: $M(\sigma|H) := \{x \in H : \sigma(x) = \inf\{\sigma(y) | y \in H\}\}$. Let $\{\phi, \phi^n\}$, $\{\psi, \psi^n\}$ be affine functions with domain $\mathbb{R}^l$ and ranges in $\mathbb{R}^m$ and $\mathbb{R}^{m'}$ respectively, such that $\phi^n \to \phi$ and $\psi^n \to \psi$ pointwize. Denote $H(\phi,\psi) := \{x \in \mathbb{R}^l \,|\phi(x) \leq 0, \text{and } \psi(x) = 0\}$.

Lemma 3 *([4] Thm I.2.2.) Assume $lim_{n\to\infty} H(\phi^n, \psi^n) = H$ for some set H. Let σ and $\{\sigma_n\}$ be linear functions such that $\sigma_n \to \sigma$ pointwise. Then $\underline{\lim}_{n\to\infty} M(\sigma_n | H(\phi^n, \psi^n)) \subset M(\sigma|H)$*

Thus the convergence of the set of solutions of a LP depends on the convergence of the feasible set. Next we present conditions for the latter convergence. Denote $I := \{i \mid 1 \leq i \leq m,\ \phi_i(x) = 0 \text{ for all } x \in H(\phi,\psi)\}$. This set depends only on the limiting LP, and identifies those *inequality* constraints which are, in fact, equalities for *all feasible points.* Let $\phi_I := \{\phi_i\ ,\ \ i \in I\}$ and denote by $\text{rank}(\phi_I, \psi)$ the rank of the matrix whose $|I| + m'$ rows are the coefficients of the linear part of $\phi_i, i \in I$, and $\psi_j, 1 \leq j \leq m'$.

Lemma 4 *([4] Cor. II.3.4) Assume $\overline{\lim}_{n\to\infty} \text{rank}(\phi^{\text{n}}_{\text{I}}, \psi^{\text{n}}) \leq \text{rank}(\phi_{\text{I}}, \psi)$.*
Then either $\lim_{n\to\infty} H(\phi^n, \psi^n) = H(\phi,\psi)$ or $H(\phi^n,\psi^n)$ is empty i.o.

3.3 Dynamic Programming

Below are some Lemmas needed later, based on well known Dynamic Programming techniques.

Lemma 5 *Let $0 \leq \beta < 1$. (i) Let $g \in U(D)$ and assume that $\lambda \in \mathbb{R}^{|\mathbf{X}|}$ satisfies*

$$\lambda_y = c(y, g(y)) + \beta \sum_{v \in \mathbf{X}} P^g_{yv} \lambda_v \tag{6}$$

Then $\lambda_y = (1-\beta)^{-1} C_\beta(y, g)$. Moreover, λ (and $C_\beta(y, g)$) are linear in c.
(ii) Let $g \in U(D)$ and assume that $\lambda \in \mathbb{R}^{|\mathbf{X}|}$ and $\eta \in \mathbb{R}$ satisfy

$$\lambda_y + \eta = c(y, g(y)) + \beta \sum_{v \in \mathbf{X}} P^g_{yv} \lambda_v \qquad \forall y \in \mathbf{X}. \tag{7}$$

Then $\lambda_y = (1-\beta)^{-1} [C_\beta(y, g) - \eta]$, or $C_\beta(y, g) = (1-\beta)\lambda_y + \eta$.
(iii) Pick some cost $c(\cdot, \cdot)$. Consider the equations with λ and η unknown:

$$\lambda_y + \eta = c(y, a) + \beta \sum_{v \in \mathbf{X}} P_{yav} \lambda_v \qquad \forall y \in \mathbf{X}, \forall a \in \mathbf{A}. \tag{8}$$

Then (8) is solvable iff $C_\beta(g, y)$ does not depend on g for any y, and the set of solutions is given by $\lambda_y = (1-\beta)^{-1} [C_\beta(y, g) - \eta]$, and η is arbitrary.

Proof: $(I - \beta P^g)$ is invertible, hence $\lambda = (I - \beta P^g)^{-1} c$ is the unique solution of (6) and is linear in c. $(1-\beta)^{-1} C_\beta(\cdot, g)$ satisfies (6), hence by the uniqueness of the solution, $\lambda_y = (1-\beta)^{-1}[C_\beta(y, g)]$. (ii) follows from (i) by considering in (6) the cost $c(y, g(y)) - \eta$ instead of $c(y, g(y))$. From (8) it follows that (7) holds for all $g \in U(D)$ with the same λ and η, so (iii) follows from (ii). ∎

Lemma 6 *Assume A1. (i) Let $g \in U(D)$ and assume that $\lambda \in \mathbb{R}^{|\mathbf{X}|}$ and $\eta \in \mathbb{R}$ satisfy $\lambda_y + \eta = c(y, g(y)) + \sum_{v \in \mathbf{X}} P^g_{yv} \lambda_v$, $\forall y \in \mathbf{X}$. Then $\eta = C_1(y, g)$ and does not depend on y, and $H(x, g) = \lambda_x + w$, where w is an arbitrary constant. η and $H(x, g)$ are linear in c. (ii) Pick some cost $c(\cdot, \cdot)$. Equation (8) with $\beta = 1$ is solvable iff $C_1(x, g)$ and $H(x, g)$ do not depend on g. The set of solutions is given by $\eta = C_1(y, g)$, and $H(x, g) = \lambda_x + w$ where w is an arbitrary constant.*

Proof: (i) is well known, and (ii) then follows as the similar claim from the previous Lemma.

4 Conditions for stability of CMDPs

We consider below both equality and inequality constraints in COP. Let $\underline{K} \stackrel{\text{def}}{=} \{1, 2, ..., K\}$. Introduce the following assumptions for a given COP:

A2: There exists $g \in U(S)$ feasible to COP such that $p^g_{a|y} > 0$ for all y, a; moreover, either $\gamma(y) > 0$ for all y or A1 holds.

A3: There exist constants $Q_1, ..., Q_K$, $k \in \underline{K}$ with not all Q_k's zero such that for any state y (i) $\sum_{k\in\underline{K}} Q_k D^k_\beta(y, u)$ does not depend on u (ii) $\sum_{k\in\underline{K}} Q_k D^k_\beta(\gamma, u) = \sum_{k\in\underline{K}} Q_k V_k$. (iii) If $\beta = 1$ then $\sum_{k\in\underline{K}} Q_k H(y, u)$ does not depend on u, $u \in U(S)$.

As we show in Thm. 7, a sufficient condition for stability is that A2 holds and A3 does not hold. Later we study the singular perturbation case, where A3 holds. A sufficient condition for A3 (ii) not to hold is the Slater condition, i.e. a policy u exists such that $D(\gamma, u) < V$. Moreover, in A3(i) it suffices to restrict to policies in $U(D)$. Indeed, suppose that A3(i) holds for all $u \in U(D)$. If there were some $v \notin U(D)$ for which the sum in A3(i) is different than for any policy in $U(D)$, then one could construct a non constrained MDP for which no policy in $U(D)$ is optimal, which is impossible in the finite case.

Remark: From Lemma 5(i), $\sum_{k=1}^K Q_k D^k_\beta(y, g)$ is equal to the cost related to policy g and an immediate cost given by $\sum_{k=1}^K Q_k d^k(\cdot, \cdot)$. Similarly, $\sum_{k=1}^K Q_k H(y, g)$ is equal to the relative cost related to policy g and an immediate cost $\sum_{k=1}^K Q_k d^k(\cdot, \cdot)$.

If COP has a single constraint then A3 becomes: for any state y, (i) $D_\beta(y, u)$ does not depend on u (ii) $D_\beta(\gamma, u) = V$. (iii) If $\beta = 1$ then $H(y, u)$ does not depend on u, $u \in U(S)$.

We say that a sequence COP_n converge to COP if the respective immediate costs, discount factors and transition probabilities converge.

Theorem 7 *Assume A2 and that A3 does not hold. Let $COP_n \to COP$ and assume that COP_n are feasible. If (1) $\beta < 1$, or (2) $\beta = 1$ and A1 holds, then (i) the optimal values of COP_n converge to the optimal value of COP (ii) if $g_n \in U(S)$ is optimal for COP_n then any limit point g of g_n is optimal for COP.*

Proof: We consider COPs that are solved through LP_β, which applies to both case (1) and (2) of the theorem. For (2) it is possible that if COP_n is obtained by perturbation of $\{P_{xay}\}$, then A1 does not hold for COP_n, and therefore if $\beta = 1$, we cannot compute the optimal value and policy for COP_n by LP_1. However, for COP_n close enough to COP, for any $u \in U(S)$, the recurrent classes of COP are subset of COP_n's, hence the optimal value and solution of COP_n is also computed through LP_β. Thus without loss of generality we assume below that COP_n are all computed through LP_β.

It follows from Lemma 2 and the remark below it that the rank of the constraint matrix A that corresponds to COP or to COP_n is at most $|\mathbf{X}| + K$. We show below that the assumptions of the theorem imply that the rank of A (corresponding to the constraint matrix of LP_β obtained from COP without the positivity constraints) is in fact equal to $|\mathbf{X}|+K$, from which it follows by Lemma 3 and 4 that the optimal values and solutions of LP_β are stable. Note that it follows from Lemma 1 that A2 ensures that there exists a feasible z such that $z>0$. Hence (ϕ_I,ψ) (in Lemma 4) represent indeed only rows of the matrix A (and do not include extra rows from constraint (5)). Since the optimal values and solutions of LP_β are stable, the proof is established. Let Y $(|\mathbf{X}|+K+1)$ be a vector, $Y=\left(Y^{(1)},Y^{(2)},Y^{(3)}\right)^T$, where $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$ are of sizes $|\mathbf{X}|$, K and 1 respectively. Consider the equation $Y^TA=0$ (A is defined below Lemma 2). We obtain

$$Y_y^{(1)}+Y^{(3)}=-\sum_{k=1}^{K}Y_k^{(2)}d^k(y,a)+\beta\sum_{v\in\mathbf{X}}Y_v^{(1)}P_{yav}\qquad\forall y\in\mathbf{X},\quad\forall a\in\mathbf{A}\tag{9}$$

The set $\mathbf{Y}$ of Y satisfying (9) is a vector space. rank $A=|\mathbf{X}|+K$ iff the rank $\mathbf{Y}=1$. For $\beta<1$, since A3 does not hold, it follows from Lemma 5 (iii) that $Y_k^{(2)}=0$. From Lemma 6 (ii) this holds also for $\beta=1$. Hence rank $\mathbf{Y}=1$, and it has as basis the vector: $Y^T=\{Y^{(1)},Y^{(2)},Y^{(3)}\}=\{(1,...,1),(0,...,0),\beta-1\}$ ■

Remark: Under arbitrary ergodic structure, if there are no constraints and $\beta<1$ then (i) and (ii) of Thm. 7 hold, since the assumptions of the theorem hold trivially.

In [2] different "Slater type" stability conditions are presented. They imply that $I\neq\emptyset$, where I, defined below Lemma 3, is the set of inequality constraints in LP_β that are achieved with equality for all z that satisfy (2)-(5). The stability then follows from the fact that the matrix of coefficients of the remaining constraints has rank $|\mathbf{X}|$. The new conditions in Thm. 7 imply stability in a different way: they ensure that the rank of the matrix of coefficients of *all constraints* in LP_β equals $|\mathbf{X}|+K$. The two approaches can be combined and a sufficient condition for stability would be that A2 holds and (i) for some $K'\subset\{1,...,K\}$, $\exists\ u\in U$ such that $D^k(\gamma,u)<V_k$ for all $k\in K'$, (ii) A3 does not hold when replacing the set $\underline{K}$ in A3 by the set $\underline{K}-K'$.

The assumption in Thm. 7 that A3 does not hold is sufficient but not necessary. Consider, e.g., the trivial case that the immediate costs $d^k(\cdot,\cdot)$ are constants and thus do not depend on y and a. In [1] Example 4.1, A3 holds and the optimal solution is not stable.

5 Singular perturbation of CMDPs

We consider COP with *equality constraints only*, for $\beta\in(0,1]$. For $\beta=1$, *we restrict COP to* $\hat{U}_\gamma$ (see remark below Lemma 1). We first quote results on perturbation of LPs and construct almost optimal solutions for the perturbed LP. We then study the singular perturbations of the discount factor and transition probabilities. We assume below A3. Assume that the dimension of the subspace spanned by the vectors Q satisfying A3 is equal to r. Assume that the columns of the matrix $\underline{Q}(K \times r)$ constitute a basis of this subspace; denote its jth column be $\underline{Q}(j)$. Let $\underline{\tilde{D}}_\beta(|\mathbf{X}| \times r)$ be the matrix given by $\underline{\tilde{D}}_\beta(y,j) \stackrel{\text{def}}{=} \sum_{k=1}^K \underline{Q}_k(j) D_\beta^k(y,u)$. Define similarly $\underline{\tilde{H}}(|\mathbf{X}| \times r)$ by $\underline{\tilde{H}}(y,j) \stackrel{\text{def}}{=} \sum_{k=1}^K \underline{Q}_k(j) H^k(y,u)$, $u \in U(S)$. For K=1 we have in particular $\underline{\tilde{D}}_\beta(y) \stackrel{\text{def}}{=} D_\beta(y,u)$, $\underline{\tilde{H}}(y) \stackrel{\text{def}}{=} H^1(y,u)$.

5.1 Singular Perturbations in Linear Programs

$$\text{Consider the LP: Find } x \text{ that } min\{(c_0^T + \epsilon c_1^T)x | (A_0 + \epsilon A_1)x = b_0 + \epsilon b_1, x \geq 0\} \tag{10}$$

where $A_0(m \times n)$, $A_1(m \times n)$, $c_0(n \times 1)$, $c_1(n \times 1)$, $b_0(m \times 1)$, $b_1(m \times 1)$ denote matrices. Let Y be a vector satisfying $Y^T A_0 = 0$ and $Y^T b_0 = 0$. Assume that the dimension of the subspace spanned by all such vectors Y is equal to l. Assume that the columns of the matrix $\underline{Y}(m \times k)$ constitute a basis of this subspace. Consider the following modified reduced LP:

$$min\{c_0^T x | A_0 x = b_0, \underline{Y}^T A_1 x = \underline{Y}^T b_1, x \geq 0\} \tag{11}$$

Lemma 8 *([7] Thm. 3.3.) Suppose (a) $\exists \bar{x} > 0$, $A_0 \bar{x} = b_0$, $\underline{Y}^T A_1 \bar{x} = \underline{Y}^T b_1$;*
(b) *rank* $\begin{bmatrix} A_0 & b_0 \\ \underline{Y}^T A_1 & \underline{Y}^T b_1 \end{bmatrix} = m$*; (c) $\emptyset \neq \theta^*(\epsilon) \subset E$, for $0 \leq \epsilon \leq \epsilon_0$ and some $\epsilon_0 > 0$, where $\theta^*(\epsilon)$ is the set of solutions to (10) and E is a compact set in $\mathbb{R}^n$. Then, (i) The set of limiting points of the solutions to (10), as $\epsilon \to 0$, is contained in the set of solutions to (11). (ii) The optimal value of (10) converges to the optimal value of (11) as $\epsilon \to 0$.*

Note that any solution of (11) is "almost" optimal for (10) if it is feasible for (10) (for ϵ small enough). This is the case in LPs that arise in singular perturbations of *non constrained MDPs* [6, 7]. However, this simple construction of approximating solutions fails once additional constraints are involved in MDPs since it need not be feasible for (10).

Let $\bar{x}(0)$ be a basic optimal solution for (11). By some correction of order ϵ we obtain below a feasible solution for (10). The value of the objective function in (10) is then within $O(\epsilon)$ of the optimal value, and hence the new solution is almost optimal for (10). It follows from Lemma 8 (b) that for small enough ϵ, rank$\hat{A} = m$ where $\hat{A} = \begin{bmatrix} A_0 + \epsilon A_1 & b_0 + \epsilon b_1 \\ \underline{Y}^T A_1 & \underline{Y}^T b_1 \end{bmatrix}$. Hence one may choose a set $M \subset \{1, 2, ..., k+m\}$, $|M| = m$, such that the rows of $\hat{A}$, whose indices are in M, are linearly independent. Let $\tilde{A}_0$, $\tilde{A}_1$, $\tilde{b}_0$ and $\tilde{b}_1$ be the matrices obtained by restricting only to rows whose indices are in M of $\begin{bmatrix} A_0 \\ \underline{Y}^T A_1 \end{bmatrix}$, $\begin{bmatrix} A_1 \\ 0(k \times n) \end{bmatrix}$, $\begin{bmatrix} b_0 \\ \underline{Y}^T b_1 \end{bmatrix}$ and $\begin{bmatrix} b_1 \\ 0(k \times 1) \end{bmatrix}$, respectively ($0(r \times s)$ means a zero matrix of size $r \times s$). It follows that (10) is equivalent to

$$min\{(c_0^T + \epsilon c_1^T)x | (\tilde{A}_0 + \epsilon \tilde{A}_1)x = \tilde{b}_0 + \epsilon \tilde{b}_1, x \geq 0\} \tag{12}$$

Assume that the basic solution $\bar{x}(0)$ has m strictly positive entries. Let A^B denote the matrix composed of the columns of a matrix A that correspond to that basis. It follows that $\tilde{A}_0^B$ is invertible. To obtain an almost optimal solution $\hat{x}$ for (10) we calculate the coefficiences $\bar{x}(i)$ in the expansion $\hat{x} = \sum_{i=0}^{\infty} \epsilon^i \bar{x}(i)$ recursively by substituting $\hat{x}$ into (12) and then equating for same powers of ϵ. As first order approximation we obtain $\hat{x}^B = \bar{x}^B(0) + \left[\tilde{A}_0^B\right]^{-1} (\tilde{b}_1 - \tilde{A}_1^B \bar{x}^B(0))$, and the other entries of $\hat{x}$ are zero. (Calculation of higher terms of the expansion involve only multiplications and summations, and no further inversions of matrices). Since $\bar{x}(0)$ has m strictly positive entries, it follows that for ϵ small enough $\hat{x}$ satisfies also the positivity constraints.

5.2 Singular perturbations of the discount factor

Introduce the modified reduced LP for $\beta < 1$, $\mathbf{LPr}_\beta$: find $z^* := \{z^*(y,a)\}_{y,a}$ that minimize $c \cdot z$ s.t.

$$\sum_{y,a} z(y,a) \left[\delta_v(y) - \beta P_{yav}\right] = [1-\beta]\gamma(v) \qquad v \in \mathbf{X} \tag{13}$$

$$d^k \cdot z = V_k \quad 1 \leq k \leq K, \qquad \sum_{y,a} z(y,a) = 1, \qquad z \geq 0 \tag{14}$$

$$\sum_{y,a} \left[\sum_{v \in \mathbf{X}} \underline{\tilde{D}}_\beta(v,s) P_{yav}\right] z(y,a) = \sum_{v \in \mathbf{X}} \gamma(v) \underline{\tilde{D}}_\beta(v,s) \qquad s = 1, ..., r \tag{15}$$

The respective LP for $\beta = 1$ is obtained by substituting $\beta = 1$ in (13) and replacing (15) with

$$\sum_{y,a} \left[\sum_{v \in \mathbf{X}} \underline{\tilde{H}}(v,s) P_{yav}\right] z(y,a) = \sum_{v \in \mathbf{X}} \gamma(v) \underline{\tilde{H}}(v,s) \qquad s = 1, ..., r$$

We treat below the singular perturbations of COP_β with respect to β. Below, A and b (defined below Lemma 2) replace A_0 and b_0 used in Lemma 8. We first identify the matrix $A_1(\{|\mathbf{X}| + K + 1\} \times \{|\mathbf{X}| \times |\mathbf{A}|\})$ and $b_1(\{|\mathbf{X}| + K + 1\} \times 1)$ that appears in Lemma 8. With the decomposition of A_1 (and b_1) similar to the decomposition of A, i.e. $A_1 = \left(A_1^{(1)}, A_1^{(2)}, A_1^{(3)}\right)^T$, we have $A_1^{(1)}(v, \{y,a\}) = -P_{yav}$, $b_1^{(1)}(v) = -\gamma(v)$, and all other entries of A_1 and b_1 are zero.

Next we identify the matrix $\underline{Y}$ from Lemma 8. It should satisfy both $\underline{Y}^T A = 0$ and $\underline{Y}^T b = 0$. Recall that the equality $Y^T A = 0$ yields (9). It then follows from A3 and Lemma 5 (iii) that $\underline{Y}^{(2)}$

has the same number of columns as $\underline{Q}$, and we may choose $\underline{Y}_k^{(2)}(j) = -\underline{Q}_k(j)$. Let $\beta < 1$. It follows from Lemma 5 (iii) that $\underline{Y}^{(3)}(j)$ may be chosen arbitrarily and

$$\underline{Y}_y^{(1)}(j) = (1-\beta)^{-1}\left(\underline{\tilde{D}}_\beta(y,j) - \underline{Y}^{(3)}(j)\right) \tag{16}$$

The equality $\underline{Y}^T b = 0$ yields

$$(1-\beta)\sum_y \gamma(y)\underline{Y}_y^{(1)}(j) + \sum_{k=1}^{K} \underline{Y}_k^{(2)}(j)V_k + \underline{Y}^{(3)}(j) = 0 \tag{17}$$

from which we obtain

$$\sum_y \gamma(y)\underline{Y}_y^{(1)}(j) = -(1-\beta)^{-1}\sum_{k=1}^{K} \underline{Y}_k^{(2)}(j)V_k + \underline{Y}^{(3)}(j) \tag{18}$$

From From (16), (18) we conclude that the probabilistic meaning of $\underline{Y}^T b = 0$ is

$$\sum_{y\in\mathbf{X}} \gamma(y)\left[\sum_{k=1}^{K} \underline{Q}_k(j)D_\beta(y,u)\right] = \sum_{k=1}^{K} \underline{Q}_k(j)V_k \qquad j = 1, ..., r \tag{19}$$

which is directly implied by the constraints $D_\beta(\gamma, u) = V$ in COP.

For $\beta = 1$ we have from Lemma 6(ii) instead of (16):

$$\underline{Y}^{(3)} = \underline{\tilde{D}}_1, \qquad \underline{Y}_y^{(1)}(j) = \underline{\tilde{H}}(y,j) + w(j) \quad j = 1, ..., r,\ y \in \mathbf{X} \tag{20}$$

where $w(j)$ is an arbitrary constant that depends on j. (17) holds for $\beta = 1$ too, and yields directly (19) in the simpler form

$$-Y_y^{(3)} = \sum_{k=1}^{K} \underline{Q}_k(j)D_1(y,u) = \sum_{k=1}^{K} \underline{Q}_k(j)V_k \qquad j = 1, ..., r,\ \forall y \in \mathbf{X} \tag{21}$$

To obtain the jth additional constraint appearing in the modified reduced LP, $Y^T A_1 z = Y^T b_1$, we compute $\underline{Y}^T A_1$ and $\underline{Y}^T b_1$. These terms also appear in the matrix $A' \stackrel{\text{def}}{=} \begin{bmatrix} A & b \\ \underline{Y}^T A_1 & \underline{Y}^T b_1 \end{bmatrix}$ whose rank should be determined in order to check (b) in Lemma 8. Denote $A'^{(i)} = \left[A^{(i)}, b^{(i)}\right]$, $i = 1, 2, 3$ and let $A'^{(4)}(r \times \{|\mathbf{X}| \times |\mathbf{A}| + 1\})$ be the matrix obtained by the remaining r rows of A'. The (y,a) element of the jth row of $A'^{(4)}$ is given by $A'^{(4)}(j, (y,a)) = -\sum_v P_{yav}Y_v^{(1)}(j)$. The last element of the jth row of $A'^{(4)}$ is given by $-Y_\gamma(j)$ where $Y_\gamma(j) \stackrel{\text{def}}{=} \sum_y \gamma(y)\underline{Y}_y^{(1)}(j)$. Substituting the expressions for $Y^{(1)}$ from (16) for $\beta < 1$ or from (20) for $\beta = 1$ in the additional constraint (see Lemma 8) $\underline{Y}^T A_1 z = \underline{Y}^T b_1$, yields (15). Note that $Y^{(3)}$ (resp. $w(j)$) disappears for $\beta < 1$ (resp. $\beta = 1$) in the new constraints.

In order to check the rank of A', define $\tilde{A}(\{|\mathbf{X}| + K + 1 + r\} \times \{|\mathbf{X}| \times |\mathbf{A}| + 1\})$, $\tilde{A} = \left(\tilde{A}^{(1)}, \tilde{A}^{(2)}, \tilde{A}^{(3)}, \tilde{A}^{(4)}\right)^T$ by: $\tilde{A}^{(i)} \stackrel{\text{def}}{=} (A^{(i)}, b^{(i)})$ for $i = 1, 2, 3$; the (y,a) entry of the jth row of $\tilde{A}^{(4)}$ is $\tilde{A}^{(4)}(j, (y, a)) \stackrel{\text{def}}{=} \underline{Y}_y^{(1)}(j)$, and the last entry is $Y_\gamma(j)$.

Theorem 9 *Assume (1) A2 and A3 hold for COP_β; (2) $\exists\ \epsilon_0 > 0$ such that $COP_{\beta+\epsilon}$ is feasible for any $0 \le \epsilon \le \epsilon_0$; (3) $\mathrm{rank}\tilde{A} = |\mathbf{X}| + 1$. Then (i) The optimal value of $COP_{\beta+\epsilon}$ converges to the optimal value of LPr_β. (ii) Let Z^* be the set of optimal solutions to the LPr_β. Let U^* be the subset of $U(S)$ obtained from Z^* by $p^u_{a|y} := \frac{z(y,a)}{\sum_a z(y,a)}$, $z \in Z^*$. Then the set of all limit points of optimal stationary policies for $COP_{\beta+\epsilon}$ is contained in U^*.*

Proof: The optimal policy and value of $COP_{\beta+\epsilon}$ are obtained using $LP_{\beta+\epsilon}$ from Section 3.1 and applying Lemma 1 (3) (and the remark below the lemma). It remains to show that (a,b,c) in Lemma 8 hold with $\mathbf{LPr}_\beta$ replacing (11) and $LP_{\beta+\epsilon}$ replacing (10). A2 implies condition (a). By Lemma 1, condition (c) follows from assumption (2) of Thm. 9. Next we establish condition (b). Since $Y^T A = 0$ implies (9),

$$A'^{(4)}(j, (y, a)) = -\sum_v P_{yav} \underline{Y}_v^{(1)}(j) = -\beta^{-1}\left[\underline{Y}_v^{(1)}(j) + \underline{Y}^{(3)}(j) + \sum_{k=1}^{K} \underline{Y}_k^{(2)}(j) d^k(y, a)\right]$$

Note that for $\beta < 1$, we have from (18) and from $\underline{Y}_k^{(2)}(j) = -\underline{Q}_k(j)$

$$-Y_\gamma(j) = (1-\beta)^{-1} \sum_{k=1}^{K} (\underline{Y}_k^{(2)}(j) V_k + \underline{Y}^{(3)}(j)) = (1-\beta)^{-1} \sum_{k=1}^{K} (-\underline{Q}_k(j) V_k + \underline{Y}^{(3)}(j))$$

We show that $\mathrm{rank}\tilde{A} = \mathrm{rank}A'$. Hence (b) in Lemma 8 holds by the assumption of the theorem. Indeed, the first $|\mathbf{X}| + K + 1$ rows of $\tilde{A}$ and A' are equivalent, $\tilde{A}^{(4)}$ is obtained by the following linear combination of rows of A' (and vice versa): multiply $A'^{(4)}(j, \cdot)$ by $-\beta$, subtract $\sum_{k=1}^{K} \underline{Y}_k^{(2)}(j) A'(|\mathbf{X}| + k, \cdot)$ (the (y,a) entry of $A'(|\mathbf{X}| + k, \cdot)$ is $d^k(y, a)$ and the last entry is V_k), then further subtract $\underline{Y}^{(3)}(j)(1, \ldots, 1) = \underline{Y}^{(3)}(j) A'^{(3)}$. ■

Remark: almost optimal policies for $COP_{\beta+\epsilon}$ can be obtained by the construction below Lemma 8.

Next we consider the case of one constraint and present sufficient conditions for assumption (4) in Theorem 9. Let $\Delta(\gamma, u)$ be the cost function associated with the immediate cost $\bar{d}(y, a) = \underline{\tilde{D}}_\beta(y)$.

Lemma 10 *Assume A3 (for COP_β) and assume $\beta < 1$. Let K=1. Assume that A3 (i) does not hold for the cost function $\Delta(y,u)$. Then rank A' =rank$\tilde{A} = |\mathbf{X}| + 1$.*

Proof: Define the vector $W(\{|\mathbf{X}|+2\})$, $W = \left\{W^{(1)}, W^{(2)}, W^{(3)}\right\}^T$ where $W^{(1)}$ is of size $|\mathbf{X}|$ and $W^{(2)}$ and $W^{(3)}$ of size 1. By A3, $\tilde{A}^{(2)}$ depends linearly on other rows of $\tilde{A}$, and hence rank$\tilde{A}$ =rankA^* where $A^* \stackrel{\text{def}}{=} \left\{\tilde{A}^{(1)}, \tilde{A}^{(3)}, \tilde{A}^{(4)}\right\}^T$. Assume that the dimension of the subspace spanned by the vectors W satisfying $W^T A^* = 0$ is equal to s. Assume that the columns of the matrix $\underline{W}(\{|\mathbf{X}|+2\} \times s)$ constitute a basis of this subspace. rank$A' = |\mathbf{X}|+1$ iff s =rank$\underline{W} = 1$. We prove below that this is indeed implied by the assumptions of the theorem. The equality $\underline{W}^T A^* = 0$ implies:

$$W_y^{(1)} + W^{(2)} = -W^{(3)}\underline{Y}_y^{(1)} + \beta \sum_{v\in\mathbf{X}} W_v^{(1)} P_{yav} \qquad \forall y \in \mathbf{X}, \quad \forall a \in \mathbf{A} \tag{22}$$

$$(1-\beta)\sum_{y\in\mathbf{X}} \gamma(y) W_y^{(1)} + W^{(2)} + \underline{Y}_\gamma W^{(3)} = 0 \tag{23}$$

(22) can be rewritten as:

$$W_y^{(1)} + \left[W^{(2)} - (1-\beta)^{-1}W^{(3)}\right] = -(1-\beta)^{-1}W^{(3)}\left((1-\beta)\underline{Y}_y^{(1)} + \underline{Y}^{(3)}\right) + \beta\sum_{v\in\mathbf{X}} W_v^{(1)}P_{yav}$$

$$= -(1-\beta)^{-1}W^{(3)}\tilde{D}_\beta(y) + \beta\sum_{v\in\mathbf{X}} W_v^{(1)}P_{yav} \qquad \forall y \in \mathbf{X}, \quad \forall a \in \mathbf{A} \tag{24}$$

see (16). Since (24) holds for all (y,a) if follows from Lemma 5 (iii) that for an immediate cost $d^*(y,a) = -(1-\beta)^{-1}W^{(3)}\tilde{D}_\beta(y)$ the respective cost $D^*(y,u)$ does not depend on u. However, this contradicts the assumption of the Lemma on $\Delta(y,u)$ unless $W^{(3)} = 0$. Hence, from Lemma 5 (iii), rank$\underline{W} = 1$, and $\underline{W}^T = \{(1,...,1), 1-\beta\}$, which establishes the proof. ∎

Remark: A similar Lemma holds for $\beta = 1$; then $\Delta(\gamma,u)$ is defined as the cost function associated with the immediate cost $\bar{d}(y,a) \stackrel{\text{def}}{=} \underline{\tilde{H}}(y)$.

5.3 Singular perturbations of the transition probabilities

Let the transition probabilities be given by $\{P_{xay} + \epsilon\tilde{P}_{xay}\}$ where $\sum_y \tilde{P}_{xay} = 0$ for all x,a. Denote the respective CMDP by COP^ϵ. Introduce the modified reduced LP for $\beta < 1$:
$\tilde{\mathbf{L}}\mathbf{Pr}$: Find $z^* := \{z^*(y,a)\}_{y,a}$ that minimizes $c \cdot z$ s.t. (13),(14) and

$$\sum_{y,a}\left[\sum_{v\in\mathbf{X}} \underline{\tilde{D}}_\beta(s,v)\tilde{P}_{yav}\right] z(y,a) = 0 \qquad s = 1,...,r \tag{25}$$

The respective LP for $\beta = 1$ is obtained by replacing (25) by $\sum_{y,a}\left[\sum_{v\in\mathbf{X}} \underline{\tilde{H}}(v,s)\tilde{P}_{yav}\right] z(y,a) = 0$. With A_1 and b_1 as in Subsection 5.2 we have $A_1^{(1)}(v,\{y,a\}) = -\tilde{P}_{yav}\ y,v \in \mathbf{X},\ a \in \mathbf{A}$, $b_1^{(1)}(v) = 0\ v \in \mathbf{X}$, and all other entries of A_1 and b_1 are zero. A, b and $\underline{Y}$ are as in Subsection 5.2. The additional constraints (25) in the modified reduced LP follow easily. A' is now given by: $A'^{(i)} \stackrel{\text{def}}{=} (A^{(i)}, b^{(i)})$ for $i = 1,2,3$; the (y,a) entry of the jth row of $A'^{(4)}$ is given by $A'^{(4)}(j,(y,a)) \stackrel{\text{def}}{=} -\sum_{v\in\mathbf{X}} \tilde{P}_{yav}\underline{Y}_v^{(1)}(j)$, and the last entry is 0. We thus obtain as in subsection 5.2:

Theorem 11 *Assume (1) A2 and A3 hold for COP_β; (2) condition (a) in Lemma 8 holds (where (10) stands for $LP_{\beta+\epsilon}$ and (11) stands for $\tilde{L}Pr$); (3)* $\text{rank}A' = |\mathbf{X}|+1$; *(4) $\exists\ \epsilon_0 > 0$ such that COP^ϵ is feasible for any $0 \le \epsilon \le \epsilon_0$. Then (i) The optimal value of COP^ϵ converges to the optimal value of $\tilde{L}Pr$. (ii) Let Z^* be the set of optimal solutions to $\tilde{L}Pr$. Let U^* be the subset of $U(S)$ obtained from Z^* by $p^u_{a|y} := \frac{z(y,a)}{\sum_a z(y,a)}$, $z \in Z^*$. Then the set of all limit points of optimal stationary policies for COP^ϵ is contained in U^*.*

Remark: almost optimal policies for COP^ϵ can be obtained by the construction below Lemma 8.

Lemma 12 *Assume A3 (for COP^0) and $\beta < 1$. Let K=1. Assume that A3 (i) does not hold for the cost function $\Delta(y,u)$. Then rank $A' = |\mathbf{X}|+1$.*

References

[1] E. Altman and A. Shwartz, "Adaptive control of constrained Markov chains", *IEEE Trans. on AC*, **36**, No. 4, 1991.

[2] E. Altman and A. Shwartz, "Sensitivity of constrained Markov Decision Problems", *Annals of Oper. Res.*, **32**, 1991.

[3] E. Altman and A. Shwartz, "Markov decision problems and state-action frequencies," *SIAM J. Control and Optimization.* **29**, No. 4, 786-809, 1991

[4] G. B. Dantzig, J. Folkman and N. Shapiro, "On the continuity of the minimum set of a continuous function", *J. Math. Anal. and Applications,* **17**, 519-548, 1967.

[5] F. Delebeque and J. Quadrat, "Optimal Control of Markov Chains Admitting Strong and Weak Interactions", *Automatica*, **17**, No. 2, 281-296, 1981.

[6] M. Abbad and J. A. Filar, "Perturbation and Stability Theory for Markov Control Problems", Research Report 90-13, Department of Math., University of Maryland, USA, 1990.

[7] V. A. Gaitsgory and A. A. Pervozvanskii, “Perturbation Theory for Mathematical Programming Problems”, *JOTA*, 389-410, 1986.

[8] L. C. M. Kallenberg, *Linear Programming and Finite Markovian Control Problems*, Math. Centre Tracts 148, Amsterdam, 1983.

[9] K. W. Ross and B. Chen, ”Optimal scheduling of interactive and non interactive traffic in telecommunication systems”, *IEEE Trans. on AC,* **33** No. 3 pp. 261-267, March 1988.