



Sensibilité de l'espérance de la récompense cumulée des modèles markoviens raides

Haïscam Abdallah, Moulaye Hamza

► To cite this version:

Haïscam Abdallah, Moulaye Hamza. Sensibilité de l'espérance de la récompense cumulée des modèles markoviens raides. [Rapport de recherche] RR-3904, INRIA. 2000. <inria-00072750>

HAL Id: inria-00072750

<https://hal.inria.fr/inria-00072750>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Sensibilité de l'espérance de la récompense cumulée
des modèles markoviens raides*

Haïscam ABDALLAH, Moulaye HAMZA

N° 3904

Mars 2000

THÈMES 1 et 4



*Rapport
de recherche*

Sensibilité de l'espérance de la récompense cumulée des modèles markoviens raides

Haïscam ABDALLAH*, Moulaye HAMZA†

Thèmes 1 et 4 — Réseaux et systèmes — Simulation et optimisation
de systèmes complexes
Projets Aladin et Armor

Rapport de recherche n° 3904 — Mars 2000 — 30 pages

Résumé : Cette étude traite de la sensibilité de l'espérance de récompense cumulée des modèles markoviens raides correspondant aux systèmes informatiques hautement fiables. Il s'agit de la dérivée partielle de cette mesure transitoire cumulative par rapport à un paramètre du générateur infinitésimal tel que le taux de panne, de réparation, la probabilité de couverture. Généralement, nous sommes confrontés au problème du temps de calcul notamment lorsque le modèle markovien est raide. Nous proposons dans ce papier une nouvelle approche basée sur la méthodologie des puissances uniformisées. L'intérêt de cette approche est sa rapidité surtout lorsque la taille de l'espace d'état est raisonnable et le temps de mission est long. La complexité temporelle de cette approche est comparée à celle de l'uniformisation standard et de la méthode IRK3. L'algorithme relatif à cette dernière méthode est implanté avec un nouveau choix du pas afin d'améliorer sa vitesse d'exécution.

Mots-clés : Système informatique, processus markoviens, raideur, espérance de récompense cumulée, sensibilité, IRK3, uniformisation, puissances uniformisées.

* Projet ALADIN

† Projet ARMOR

Sensitivity Analysis of the Expected Accumulated Reward of Stiff Markov Models

Abstract: This study deals with the sensitivity computation of the expected accumulated reward of stiff Markov models. The sensitivity of this transient cumulative measure is its derivative with respect to a parameter of the infinitesimal generator such that the fault rate, repair rate, coverage probability. In order to compute this sensitivity, we are generally faced with the problem of the computation time, especially, when the Markov model is stiff. We propose a new approach based on the uniformized power methodology. The interest of this approach is that it will save computation time if the state space is not too large and the mission time is long. The time complexity of this method is compared with respect the standard uniformization and IRK3 methods. The algorithm relatively to this ODE method is implemented with a new choice of the step which accelerates its execution.

Key-words: Computing system, Markov process, stiffness, expected accumulated reward, sensitivity, IRK3, uniformization, uniformized power.

1 Introduction

Les modèles markoviens de récompense sont des outils de préférence pour modéliser le comportement des systèmes informatiques tolérant ou non les pannes. Il s'agit des modèles combinant les aspects performance et fiabilité. Ceci revient à considérer un modèle markovien (de fiabilité) et à associer à chaque état un taux de récompense représentant le niveau de performance du système dans cet état. Les transitions vers les états à faible taux de récompense sont des transitions vers les composants défaillants. Dans les systèmes réparables, les transitions vers les états à haut niveau de performance sont des transitions vers les états opérationnels. Dans la suite de cette étude, on suppose que les taux de récompense sont tous constants.

Il est intéressant pour ce type de modèles d'évaluer non uniquement les mesures transitoires instantanées, mais aussi les mesures transitoires cumulatives telles que l'espérance de récompense cumulée (ERC) sur un intervalle donné $[0, t]$; t désigne le temps de mission du système. Comme les paramètres utilisés (taux de panne, taux de recouvrement, probabilité de couverture) sont estimés, il est alors nécessaire de déterminer la sensibilité de cette espérance aux paramètres. La sensibilité de l'ERC à un paramètre se définit comme étant sa dérivée partielle par rapport à ce paramètre. Les paramètres donnant les plus grandes sensibilités nécessitent une attention particulière. Une autre utilité de la sensibilité est qu'elle peut être utilisée pour réduire le nombre de paramètres en éliminant certains d'entre eux, si cela n'a pas d'influence significative sur les résultats. L'analyse de la sensibilité de mesures transitoires est beaucoup utilisée dans la conception des systèmes informatiques tolérant (ou non) les pannes. Une de ses applications à l'analyse des multiprocesseurs est présentée dans [1]. Nous nous intéressons, dans ce rapport, à l'évaluation numérique de la sensibilité de l'ERC des modèles markoviens de récompense raides. Ceux-ci correspondent aux systèmes informatiques hautement fiables pour lesquels les taux de panne sont beaucoup plus petits que les taux de recouvrement. Le rapport taux de recouvrement au taux de panne peut aller jusqu'à 10^8 . La raideur dépend également du temps de mission du système. Ainsi, dans [2], l'indice de raideur a été défini comme étant le produit du maximum des taux de séjour dans les états et du temps de mission du système. Ce produit peut avoir un ordre de grandeur de 10^8 .

Lors de l'évaluation de la sensibilité de l'ERC, nous sommes confrontés au problème du temps de calcul des ordinateurs. Nous allons procéder au calcul de cette sensibilité par les méthodes analytiques. Les méthodes de simulation de Monte-

Carlo ont été appliquées pour estimer la sensibilité de mesures transitoires et stationnaires [3, 4]. Ces méthodes servent souvent à valider les méthodes analytiques quand celles-ci trouvent leur limite en temps de calcul. Elles sont très coûteuses lorsque l'on a affaire à des modèles raides. Les modèles acycliques correspondant aux systèmes informatiques tolérant les pannes mais non réparables ont été étudiés dans [5]. Une expression de la sensibilité de l'ERC y est donnée. L'algorithme relatif à cette approche est numériquement instable lorsque le modèle est raide.

Les méthodes de résolution des équations différentielles ordinaires (ODE : Ordinary Differential Equations) sont aussi utilisées pour calculer les mesures transitoires cumulatives [6, 7, 8]. Le système d'équations différentielles dont la résolution donne la distribution cumulative (intégrale du vecteur des probabilités d'état), donc les mesures cumulatives, est non homogène. Sa résolution ne pose pas de problèmes particuliers dans la mesure où le second membre est indépendant de t . Pour calculer la sensibilité de la distribution cumulative sur $[0, t]$ et déduire celle de l'ERC, nous sommes face à un système non homogène avec un second membre dépendant de t . Pour affronter le problème du temps de calcul croissant avec la raideur, deux méthodes ODE L-stables sont souvent utilisées pour calculer les distributions transitoires instantanée et cumulatives. Il s'agit des méthodes TR-BDF2 (Trapezoid Rule Backward Difference Formulae d'ordre 2) et IRK3 (Implicit Runge-Kutta d'ordre 3). L'application de ces deux méthodes au calcul de la sensibilité de mesures transitoires instantanées a montré que la méthode TR-BDF2 souffre du temps de calcul lorsque la raideur augmente [9, 10]. Nous retenons uniquement la méthode IRK3. Nous allons d'abord l'adapter au calcul de la sensibilité de la distribution cumulative en calculant simultanément la distribution et sa dérivée partielle par rapport à un paramètre. Ensuite, comme dans le cas de la sensibilité de mesures transitoires instantanées, nous choisissons un nouveau pas augmentant la vitesse d'exécution de cette méthode. L'inconvénient des méthodes ODE en général est qu'elles ne permettent pas un contrôle de l'erreur globale de précision. Une estimation d'une certaine **erreur locale** se fait après la résolution de systèmes linéaires par une méthode itérative. La complexité temporelle ne peut donc pas être connue à l'avance.

La technique de l'uniformisation standard (SU : Standard Uniformization) consiste à déterminer l'expression du vecteur des probabilités d'état (la distribution transitoire instantanée) sous forme d'une somme infinie d'éléments positifs. La troncature de cette somme infinie permet de majorer l'**erreur globale** à un certain rang

dépendant d'une tolérance fournie a priori par l'utilisateur. L'approche SU est aussi utilisée pour calculer la sensibilité de mesures transitoires instantanées [1, 11, 12]. L'intégration sur $[0, t]$ du vecteur des probabilités d'état puis sa dérivation par rapport à un paramètre donne respectivement la distribution cumulative et sa sensibilité. Dans tous les cas, l'erreur globale de précision est majorée par une expression simple reposant sur la troncature d'une somme infinie. Malheureusement, lorsque le temps de mission est long, le temps de calcul devient prohibitif même si le nombre des états est modéré.

Afin de remédier aux problèmes du temps de calcul de la méthode SU et du contrôle de l'erreur globale par les méthodes ODE, nous proposons une nouvelle approche basée sur la méthodologie des puissances uniformisées (UP: Uniformized Power method). Cette technique permet de calculer la sensibilité de l'ERC (et l'ERC) avec un temps de calcul acceptable pour les modèles les plus raides ayant un espace d'état raisonnable. La méthode UP comprend aussi un mécanisme efficace de contrôle de l'erreur.

Ce papier est organisé comme suit. La section suivante comprend la formulation mathématique du problème. Dans la section 3, nous présentons la méthode IRK3 avec un nouveau choix du pas. La technique SU est rappelée dans la section 4. La section 5 est consacrée à la méthode proposée (UP). Nous comparons dans la section 6 la complexité temporelle des 3 méthodes; un exemple numérique illustrera cette comparaison. La section 7 comprend la conclusion de ce travail.

2 Formulation du problème

Considérons une Chaîne de Markov à Temps Continu (CMTC) homogène $\mathbf{X} = \{X_t, t \geq 0\}$ à espace d'état $\mathbb{E} = \{1, 2, \dots, M\}$ fini. Soit $R = (r_i)$ le vecteur des taux de récompense; r_i dénote le taux de récompense associé à l'état i de \mathbf{X} . On suppose que les taux de transition dépendent d'un paramètre θ ; θ pourra être le taux de panne, de recouvrement, la probabilité de couverture. Le générateur infinitésimal (ou matrice des taux de transition) de la CMTC \mathbf{X} est noté $Q(\theta) = (q_{ij}(\theta))$. Soit $\Pi(\theta, t)$ le vecteur des probabilités d'état exprimant la distribution transitoire

instantanée. L'ERC sur l'intervalle $[0, t]$ est définie par

$$E[Y(\theta, t)] = \sum_{i=1}^M r_i L_i(\theta, t) = L(\theta, t)R \quad (1)$$

où

$$L(\theta, t) = \int_0^t \Pi(\theta, s) ds \quad (2)$$

est la distribution cumulative de \mathbf{X} sur l'intervalle $[0, t]$.

La sensibilité de $E[Y(\theta, t)]$ est sa dérivée partielle par rapport à θ . Il ressort de (1) que

$$\frac{\partial}{\partial \theta} E[Y(\theta, t)] = \frac{\partial}{\partial \theta} [L(\theta, t)R] = \left[\frac{\partial}{\partial \theta} L(\theta, t) \right] R \quad (3)$$

puisque les taux de récompense sont supposés constants. Le calcul de cette sensibilité repose alors sur celui de $\frac{\partial}{\partial \theta} L(\theta, t)$, ce vecteur est noté dans la suite $S_L(\theta, t)$.

Il est bien connu que $\Pi(\theta, t)$ est solution du système homogène d'équations différentielles du premier ordre de Chapman- Kolmogorov :

$$\frac{\partial}{\partial t} \Pi(\theta, t) = \Pi(\theta, t)Q(\theta); \quad \Pi(0) \text{ donné} \quad (4)$$

Le vecteur $\Pi(0)$ désigne la distribution initiale de \mathbf{X} , cette distribution est supposée indépendante de θ . L'expression de la solution de (4) est

$$\Pi(\theta, t) = \Pi(0)P(\theta, t) \quad (5)$$

où $P(\theta, t)$ est l'exponentielle de la matrice $Q(\theta)t$:

$$P(\theta, t) = \sum_{n=0}^{\infty} Q(\theta)^n \frac{t^n}{n!}. \quad (6)$$

Il existe deux familles de méthodes permettant le calcul de $S_L(\theta, t)$. La première famille comprend les méthodes ODE : on intègre le système (4) afin d'obtenir un autre système différentiel dont la solution est $L(\theta, t)$. La dérivation par rapport à θ de

ce dernier système différentiel conduit à un nouveau système ayant $S_L(\theta, t)$ comme solution. Nous retenons de cette famille la méthode IRK3. Il s'agit d'une méthode L-stable (voir ci-dessous) particulièrement adaptée aux CMTC raides.

Les méthodes de la seconde famille consistent à calculer l'expression de $S_L(\theta, t)$ à partir de celle de $\Pi(\theta, t)$. En effet, une fois l'expression de $\Pi(\theta, t)$ obtenue, il suffit de l'intégrer sur l'intervalle $[0, t]$ puis dériver l'expression de l'intégrale par rapport à θ . La technique SU et l'approche proposée UP font partie de la seconde famille.

3 La méthode IRK3

3.1 Introduction

Rappelons que les méthodes ODE s'appliquent en général à une équation ou à un système d'équation du type $y'(t) = f(t, y(t))$; $y(t)$ est solution de cette équation. La stratégie des méthodes numériques ODE est de subdiviser l'intervalle de solution $[0, t]$ en $\{0 = t_0, \dots, t_n = t\}$ et de calculer une solution approchée de la fonction inconnue $y(t)$ aux points t_i , $i \geq 1$. Soit $y(t_i)$ (resp. y_i) la solution exacte (resp. approchée) de l'équation différentielle en t_i . Le pas en i est défini par $h_i = t_{i+1} - t_i$. Afin d'affronter la raideur, nous considérons une méthode implicite L-stable telle que IRK3 [13], [14]. Pour plus de détails sur les méthodes ODE et leurs propriétés de L-stabilité, le lecteur peut consulter [15].

L'intégration du système (4) donne

$$\frac{\partial}{\partial t} L(\theta, t) = L(\theta, t)Q(\theta) + \Pi(0); \quad L(\theta, 0) = L(0) = 0 \quad (7)$$

La dérivation par rapport à θ de ce dernier système conduit au système suivant ayant $S_L(\theta, t)$ comme solution

$$\frac{\partial}{\partial t} S_L(\theta, t) = S_L(\theta, t)Q(\theta) + L(\theta, t) \frac{\partial}{\partial \theta} Q(\theta); \quad S_L(\theta, 0) = S_L(0) = 0 \quad (8)$$

La résolution du système (8) par une méthode ODE quelconque nécessite la connaissance préalable de $L(\theta, t)$ en t_i . L'implantation des méthodes considérées pour résoudre (7) et (8) séparément est difficile. De plus, le système (8) est non homogène avec un second membre dépendant de t . De même, le système (7) est non homogène mais avec un second membre indépendant de t . Ceci ne complique pas la solution

de ce système.

Rappelons que les méthodes L-stables s'appliquent à des systèmes homogènes du type $y' = \lambda y$. L'adaptation de la méthode IRK3 au calcul de $S_L(\theta, t)$ est alors nécessaire. Une autre possibilité offerte est la résolution des systèmes (7) et (8) simultanément en les réunissant pour obtenir le système suivant

$$\frac{\partial}{\partial t} \begin{pmatrix} S_L(\theta, t) & L(\theta, t) \end{pmatrix} = \begin{pmatrix} S_L(\theta, t) & L(\theta, t) \end{pmatrix} \begin{pmatrix} Q(\theta) & 0 \\ \frac{\partial}{\partial \theta} Q(\theta) & Q(\theta) \end{pmatrix} + \begin{pmatrix} 0 & \Pi(0) \end{pmatrix} \quad (9)$$

avec $(S_L(\theta, 0) \ L(\theta, 0)) = (0 \ 0)$.

Finalement, on transforme le système précédent afin d'obtenir un système homogène de premier ordre

$$\frac{\partial}{\partial t} V(\theta, t) = V(\theta, t) B(\theta) \quad (10)$$

où $V(\theta, t) = (S_L(\theta, t) \ L(\theta, t) \ 1)$ et

$$B(\theta) = \begin{pmatrix} Q(\theta) & 0 & 0 \\ \frac{\partial}{\partial \theta} Q(\theta) & Q(\theta) & 0 \\ 0 & \Pi(0) & 0 \end{pmatrix}$$

La condition initiale est telle que $V(\theta, 0) = (0 \ 0 \ 1)$.

L'ordre de la matrice carrée $B(\theta)$ est $2M + 1$. Généralement, $B(\theta)$ est stockée de façon compacte et l'intérêt porte sur les deux premières colonnes de $V(\theta, t)$ et $B(\theta)$. Notons en passage que le vecteur $\Pi(0)$ contient souvent un seul élément non nul du fait que dans la plupart des cas $\Pi_1(0) = 1$ et $\Pi_i(0) = 0, i \geq 2$.

3.2 Application au calcul de $S_L(\theta, t)$

L'application de la méthode IRK3 au système (10) donne le schéma itératif :

$$V_{i+1} \left[I - \frac{2}{3} h_i B(\theta) + \frac{1}{6} h_i^2 B(\theta)^2 \right] = V_i \left[I + \frac{1}{3} h_i B(\theta) \right] \quad (11)$$

La méthode IRK3 requiert essentiellement la résolution du système linéaire (11). En général, $B(\theta)$ est stocké de façon compacte et une méthode itérative telle que celle de

Gauss-Seidel est utilisée pour résoudre les systèmes linéaires. la complexité temporelle dépend donc du nombre de pas, noté p , et du nombre d'itérations par pas, dont la moyenne est notée \bar{I} . Si η' désigne le nombre d'éléments non nuls de la matrice $B(\theta)^2$, la complexité temporelle de la méthode IRK3 est en $O(\bar{I}p\eta')$.

Étudions maintenant la précision de la méthode IRK3. Appliquée à l'équation (10), cette méthode calcule $V(t)$ avec un vecteur d'erreur locale $\varepsilon(h_i)$ commise au pas h_i (donc au temps $t + h_i$). On calcule ensuite une *LTE* (erreur locale de troncature) comme la norme (par exemple la norme infinie) du vecteur $\varepsilon(h_i)$. L'expression de $\varepsilon(h_i)$ est la suivante [16] :

$$\varepsilon(h_i) = \frac{h_i^4}{72}V(t)B(\theta)^4 + O(h_i^5) \quad (12)$$

Le calcul direct de $\varepsilon(h_i)$ ne pose pas de problème du fait que $B(\theta)^2$ et $V(t)B(\theta)$ ont déjà été calculés pour former les deux membres de l'équation (11). Afin d'obtenir $V(t)B(\theta)^4$, il suffit de faire deux produits matrice-vecteur, l'un avec la matrice $B(\theta)$ et l'autre avec la matrice $B(\theta)^2$. Il existe d'autres techniques d'estimation d'erreur telle que l'extrapolation de Richardson. Celle-ci permet d'obtenir une estimation de $\varepsilon(h_i)$ à l'aide de la formule suivante :

$$\varepsilon(h_i) = \frac{8}{7}[\hat{V}_1(t + h_i) - \hat{V}_2(t + h_i)]$$

où $\hat{V}_1(t + h_i)$ et $\hat{V}_2(t + h_i)$ sont obtenus à partir de $V(t)$, avec des pas respectifs de h_i et $2h_i$. Les expériences montrent que l'erreur calculée par les deux procédés a le même ordre de grandeur. L'inconvénient du deuxième procédé est l'augmentation de la complexité temporelle de IRK3. Le calcul direct par la relation (12) est adopté dans la suite.

3.3 Choix du pas et implémentation

On part d'un temps initial t_0 (ici égal à 0) et on se fixe un pas initial h_0 . La méthode IRK3 est appliquée pour obtenir V_1 , la solution en $t_1 = t_0 + h_0$. Cette méthode est ensuite appliquée en t_1 avec un nouveau pas h_1 et avec V_1 comme vecteur initial. Le résultat est V_2 . On continue ainsi jusqu'à atteindre le temps de mission t auquel la solution est désirée.

Le passage du pas i au pas $i + 1$ ne peut se faire qu'après adoption du pas h_i et résolution d'un système linéaire. Le pas est calculé en fonction de la LTE et de la tolérance locale τ fournie par l'utilisateur ($LTE \leq \tau$) à l'aide de la formule suivante :

$$h_i \leq h_{i-1} \left(\frac{\tau}{\|\varepsilon(h_i)\|} \right)^{\frac{1}{r+1}}$$

où r désigne l'ordre d'une méthode ODE (ici, $r = 3$). De plus, afin d'éviter d'une part un grand nombre de pas (si h est trop petit) et d'autre part, une mauvaise précision (si h est trop grand), on fixe des limites h_{min} et h_{max} telles que $h_{min} \leq h_i \leq h_{max}$. Un exemple de choix de $h_{min} = 10^{-7}$ et $h_{max} = 10$ est donné dans [7]. Pour éviter une succession de pas faibles (donc une augmentation du temps de calcul), nous avons d'abord calculé l'erreur et ensuite choisi le pas optimal par la formule de relaxation précédente. Ce pas est **le plus grand h pour lequel l'erreur reste inférieure à la tolérance locale fournie τ** . Cette technique nous a permis, en partant d'un pas initial $h_0 = 10^{-7}$, d'avoir des valeurs du pas relativement grandes (voir section 6).

Le contrôle de l'erreur globale de précision par les méthodes ODE est difficile. En conséquence, la LTE ne garantit pas toujours la précision désirée par l'utilisateur. De plus, comme les nombres de pas et d'itérations par pas ne sont pas connus à l'avance, on ne peut alors pas estimer le temps de calcul pour des valeurs données de t et M à partir de la complexité temporelle. Il est important de noter que la complexité temporelle ne croît pas forcément linéairement avec t du fait que le pas est choisi en fonction de la LTE (variable) et de la tolérance locale fournie par l'utilisateur.

La technique SU que nous allons présenter calcule la sensibilité avec une erreur obtenue par troncature d'une somme infinie. Le mécanisme de contrôle de l'erreur globale par cette méthode est assez efficace.

4 La méthode SU

La méthode SU [12], [2], [8], [17], associe à la CMTC d'origine une Chaîne de Markov à Temps Discret (CMTD) dont les transitions entre états sont régies par un processus de Poisson. Ainsi, à la CMTC \mathbf{X} gouvernée par $\Pi(0)$ et $Q(\theta)$, on

associera une CMTD ayant aussi $\Pi(0)$ comme distribution initiale mais une matrice des probabilités de transition $\tilde{P}(\theta)$ telle que

$$\tilde{P}(\theta) = I + Q(\theta)/q ; q \geq \max_i |q_{ii}(\theta)| \quad (13)$$

où I est la matrice identité d'ordre M . Il s'ensuit que $Q(\theta) = q(\tilde{P}(\theta) - I)$ et que

$$P(\theta, t) = e^{qt\tilde{P}(\theta) - qtI} = e^{-qt} I e^{qt\tilde{P}(\theta)} = e^{-qt} e^{qt\tilde{P}(\theta)} \quad (14)$$

L'expression de $P(\theta, t)$ est donc :

$$P(\theta, t) = \sum_{n=0}^{\infty} p(n, qt) \tilde{P}(\theta)^n \quad (15)$$

où

$$p(n, qt) = e^{-qt} \frac{(qt)^n}{n!} \quad (16)$$

est la distribution de Poisson de moyenne qt ; q s'appelle le taux uniformisé. En définissant le vecteur $\tilde{\Pi}^{(n)}(\theta)$ par $\tilde{\Pi}^{(n)}(\theta) = \Pi(0)\tilde{P}(\theta)^n$, on obtient à partir de (5) et (15)

$$\Pi(\theta, t) = \sum_{n=0}^{\infty} p(n, qt) \tilde{\Pi}^{(n)}(\theta). \quad (17)$$

Les vecteurs $\tilde{\Pi}^{(n)}(\theta)$ sont tels que

$$\tilde{\Pi}^{(n)}(\theta) = \tilde{\Pi}^{(n-1)}(\theta)\tilde{P}(\theta) \text{ pour } n \geq 1 \text{ avec } \tilde{\Pi}^{(0)}(\theta) = \Pi(0).$$

Dans la pratique, on calcule une valeur approchée des probabilités d'état à partir de l'équation (17) en tronquant la somme à un certain terme N_T , de sorte que l'erreur de troncature soit bornée. On choisit N_T tel que

$$\varepsilon \geq 1 - \sum_{n=0}^{N_T} p(n, qt). \quad (18)$$

où ε est une tolérance spécifiée a priori par l'utilisateur. Le calcul de $\Pi(\theta, t)$ requiert un produit vecteur-matrice par itération. Lorsque l'indice de raideur qt augmente, le temps de calcul devient prohibitif puisque, d'après l'inégalité (18), $N_T > qt$. Afin d'assurer l'existence de la sensibilité de mesures transitoires, on suppose que $q > \max_i |q_{ii}(\theta)|$.

Nous avons vu (relation (3)) que le calcul de $\frac{\partial}{\partial \theta} E[Y(\theta, t)]$ repose sur celui de $S_L(\theta, t)$. Déterminons d'abord l'expression de $L(\theta, t)$. L'intégration de la relation (17) donne :

$$L(\theta, t) = \sum_{n=0}^{\infty} \tilde{\Pi}^{(n)}(\theta) \int_0^t p(n, qs) ds \quad (19)$$

On peut vérifier aisément que pour tout $n \in \mathbf{IN}$,

$$\int_0^t p(n, qs) ds = \frac{1}{q} \sum_{k=n+1}^{\infty} p(k, qt)$$

En remplaçant ce résultat dans l'expression (19), on obtient

$$\begin{aligned} L(\theta, t) &= \frac{1}{q} \sum_{n=0}^{\infty} \tilde{\Pi}^{(n)}(\theta) \sum_{k=n+1}^{\infty} p(k, qt) \\ &= \frac{1}{q} \sum_{n=1}^{\infty} p(n, qt) \sum_{k=0}^{\infty} \tilde{\Pi}^{(k)}(\theta) \\ &= t \sum_{n=1}^{\infty} p(n, qt) \frac{1}{n+1} \sum_{k=0}^{n-1} \tilde{\Pi}^{(k)}(\theta) \end{aligned}$$

Finalement,

$$L(\theta, t) = t \sum_{n=0}^{\infty} p(n, qt) \frac{1}{n+1} \sum_{k=0}^n \tilde{\Pi}^{(k)}(\theta) \quad (20)$$

Soit ε_L une tolérance donnée et N_L le rang de troncature de la somme infinie précédente. Nous pouvons écrire :

$$L(\theta, t) = t \sum_{n=0}^{N_L} p(n, qt) \frac{1}{n+1} \sum_{k=0}^n \tilde{\Pi}^{(k)}(\theta) + E_L$$

La norme infinie du vecteur d'erreur E_L vérifie

$$\|E_L\|_\infty \leq t \sum_{n=N_L+1}^{\infty} p(n,qt) \frac{1}{n+1} \sum_{k=0}^n \left\| \tilde{\Pi}^{(k)}(\theta) \right\|_\infty$$

Comme pour tout $k \in \mathbb{N}$, le vecteur $\frac{1}{n+1} \sum_{k=0}^n \tilde{\Pi}^{(k)}(\theta)$ est stochastique, on a :

$$\|E_L\|_\infty \leq t \sum_{n=N_L+1}^{\infty} p(n,qt) = t \left[1 - \sum_{n=0}^{N_L} p(n,qt) \right]$$

L'erreur maximale commise sur $E[Y(\theta,t)]$, notée E_Y , est telle que

$$E_Y \leq t \left[1 - \sum_{n=0}^{N_L} p(n,qt) \right] \|R\|_\infty \leq \varepsilon_L$$

L'erreur de troncature commise et la complexité temporelle de la méthode SU dépendent de la façon avec laquelle cette troncature est effectuée. Une première possibilité est de considérer $N_T = N_L$. D'après l'inégalité (18), on a :

$$E_Y \leq t\varepsilon \|R\|_\infty \quad (21)$$

L'autre possibilité offerte est le calcul de $E[Y(\theta,t)]$ avec une erreur absolue maximale ε_L . Ceci revient à prendre

$$\varepsilon = \frac{\varepsilon_L}{t \|R\|_\infty}$$

Il est clair que $\varepsilon < \varepsilon_L$ et que l'écart devient très important lorsque le temps de mission t augmente. La valeur de N_L risque alors d'être très grande par rapport à celle de qt . Ceci risque d'augmenter considérablement la complexité temporelle de la méthode SU. Pour éviter ces difficultés, nous retenons $N_L = N_T$.

4.1 Calcul de $S_L(\theta,t)$

Afin d'obtenir l'expression de $S_L(\theta,t)$, nous allons dériver l'équation (20) par rapport à θ , ce qui donne

$$S_L(\theta,t) = t \sum_{n=0}^{\infty} p(n,qt) \frac{1}{n+1} \sum_{k=0}^n \frac{\partial}{\partial \theta} \tilde{\Pi}^{(k)}(\theta) \quad (22)$$

Les vecteurs $\frac{\partial}{\partial \theta} \tilde{\Pi}^{(k)}(\theta)$ vérifient

$$\frac{\partial}{\partial \theta} \tilde{\Pi}^{(k)}(\theta) = \left[\frac{\partial}{\partial \theta} \tilde{\Pi}^{(k-1)}(\theta) \right] \tilde{P}(\theta) + \tilde{\Pi}^{(k-1)}(\theta) \left[\frac{\partial}{\partial \theta} \tilde{P}(\theta) \right], k \geq 1 \quad (23)$$

La condition initiale est $\frac{\partial}{\partial \theta} \tilde{\Pi}^{(0)}(\theta) = 0$ car $\tilde{\Pi}^{(0)}(\theta) = \Pi(0)$.

Dans la pratique, on calcule simultanément $L(\theta, t)$ et $S_L(\theta, t)$ en tronquant les deux sommes infinies (20) et (22) au même rang N_L . Les raisons invoquées ci-dessus relativement à la complexité temporelle nous amènent à considérer N_T comme terme de troncature des différentes sommes infinies données par la méthode SU. Le vecteur d'erreur maximale commise sur $S_L(\theta, t)$, notée F_L , satisfait alors :

$$\|F_L\|_\infty \leq t \sum_{n=N_T+1}^{\infty} p(n, qt) \frac{1}{n+1} \sum_{k=0}^n \left\| \frac{\partial}{\partial \theta} \tilde{\Pi}^{(k)}(\theta) \right\|_\infty$$

D'après la relation (23), on peut démontrer par récurrence que pour tout $k \in \mathbf{IN}$,

$$\left\| \frac{\partial}{\partial \theta} \tilde{\Pi}^{(k)}(\theta) \right\|_\infty \leq k \left\| \frac{\partial}{\partial \theta} \tilde{P}(\theta) \right\|_\infty$$

Il s'ensuit que

$$\begin{aligned} \|F_L\|_\infty &\leq t \sum_{n=N_T+1}^{\infty} p(n, qt) \frac{1}{n+1} \sum_{k=0}^n k \left\| \frac{\partial}{\partial \theta} \tilde{P}(\theta) \right\|_\infty \\ &= t \sum_{n=N_T+1}^{\infty} p(n, qt) \frac{1}{n+1} \frac{n(n+1)}{2} \left\| \frac{\partial}{\partial \theta} \tilde{P}(\theta) \right\|_\infty \\ &= \frac{t}{2} \left\| \frac{\partial}{\partial \theta} \tilde{P}(\theta) \right\|_\infty \sum_{n=N_T+1}^{\infty} np(n, qt) \\ &= \frac{t}{2} \left\| \frac{\partial}{\partial \theta} \tilde{P}(\theta) \right\|_\infty qt \sum_{n=N_T}^{\infty} p(n, qt) \\ &= \frac{qt^2}{2} \left\| \frac{\partial}{\partial \theta} \tilde{P}(\theta) \right\|_\infty \sum_{n=N_T}^{\infty} p(n, qt) \\ &= \frac{t^2}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \sum_{n=N_T}^{\infty} p(n, qt) \end{aligned}$$

En tenant compte de l'inégalité (18), l'erreur maximale commise sur $\frac{\partial}{\partial \theta} E[Y(\theta, t)]$, notée F_Y , satisfait alors :

$$F_Y \leq \frac{t^2}{2} \|R\|_\infty \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty [\varepsilon + p(N_T, qt)] \quad (24)$$

Le calcul de $\frac{\partial}{\partial \theta} E[Y(\theta, t)]$ requiert essentiellement 3 produits vecteur-matrice par itération. La complexité temporelle de la méthode SU est donc en $O(3N_T M^2)$. En utilisant un stockage compact de $Q(\theta)$ et de sa dérivée $\frac{\partial}{\partial \theta} Q(\theta)$, la complexité temporelle pourra être réduite à $O(N_T(2\eta + \eta_s))$.

Pour une tolérance donnée, la valeur de N_T reste toujours supérieure à celle de qt . Lorsque la raideur (donc qt) augmente, le temps de calcul de SU devient prohibitif. L'approche que nous proposons traite efficacement de telle classe de problèmes.

5 L'approche UP

Cette approche a déjà été utilisée pour calculer plusieurs mesures transitoires et la sensibilité de $\Pi(\theta, t)$ [9, 10, 17, 18, 19]. Elle est basée sur le choix d'un nombre entier positif m tel que $qt_0 < 1$ où $t_0 = t/2^m$. Ce choix est justifié plus tard. D'un point de vue pratique, si par exemple $qt = 10^8$ et $m = 30$ alors $qt_0 < 1$. Soit $C(\theta, t)$ la matrice définie par

$$C(\theta, t) = \frac{1}{t} \int_0^t P(\theta, s) ds \quad (25)$$

Notons que la matrice $C(\theta, t)$ est stochastique et que

$$L(\theta, t) = t\Pi(0)C(\theta, t) \quad (26)$$

En conséquence, nous avons

$$E[Y(\theta, t)] = t\Pi(0)C(\theta, t)R \quad (27)$$

On en déduit la sensibilité de $E[Y(\theta, t)]$, notée $S_Y(\theta, t)$, en dérivant la relation précédente par rapport à θ :

$$S_Y(\theta, t) = t\Pi(0) \left[\frac{\partial}{\partial \theta} C(\theta, t) \right] R \quad (28)$$

Soit $D_C(\theta, t)$ la matrice désignant $\frac{\partial}{\partial \theta} C(\theta, t)$. La technique UP consiste à calculer dans un premier temps la matrice $D_C(\theta, t_0)$ à l'aide de la méthode SU. Dans un deuxième temps, une nouvelle relation de récurrence permet le retour à $D_C(\theta, t) = D_C(\theta, 2^m t_0)$ en partant de $D_C(\theta, t_0)$. La sensibilité de $E[Y(\theta, t)]$ se déduit par la simple relation (28). En fait, avec cette relation, on calcule une suite de $m + 1$ valeurs $S_Y(\theta, t_l)$, $l = 0, 1, \dots, m$ où $t_l = 2^l t_0$.

Construisons d'abord cette relation de récurrence. On peut démontrer facilement que pour tout $l \in \mathbb{N}$,

$$C(\theta, t_l) = \left[\frac{P(\theta, t_{l-1}) + I}{2} \right] C(\theta, t_{l-1}) \quad (29)$$

La dérivation par rapport à θ de la relation précédente donne

$$D_C(\theta, t_l) = \frac{1}{2} [D_P(\theta, t_{l-1})] C(\theta, t_{l-1}) + \frac{1}{2} [P(\theta, t_{l-1}) + I] D_C(\theta, t_{l-1}) \quad (30)$$

où $D_P(\theta, t_{l-1})$ désigne la matrice $\frac{\partial}{\partial \theta} P(\theta, t_{l-1})$. Les matrices $P(\theta, t_{l-1})$ s'obtiennent à partir des équations de Chapman-Kolmogorov :

$$P(\theta, t_{l-1}) = [P(\theta, t_{l-2})]^2 \quad (31)$$

Les matrices dérivées $D_P(\theta, t_{l-1})$ se déduisent par dérivation par rapport à θ des équations précédentes :

$$D_P(\theta, t_{l-1}) = D_P(\theta, t_{l-2})P(\theta, t_{l-2}) + P(\theta, t_{l-2})D_P(\theta, t_{l-2}) \quad (32)$$

L'expression de $P(\theta, t_0)$ par la méthode SU est donnée par la formule (15). L'intégration de cette relation se fait de la même façon que dans (20), ce qui donne

$$C(\theta, t_0) = \sum_{n=0}^{\infty} p(n, qt_0) \frac{1}{n+1} \sum_{k=0}^n \tilde{P}^k(\theta) \quad (33)$$

Les dérivées respectives de $P(\theta, t_0)$ et $C(\theta, t_0)$ sont :

$$D_P(\theta, t_0) = \sum_{n=0}^{\infty} p(n, qt_0) \frac{\partial}{\partial \theta} \tilde{P}^n(\theta) \quad (34)$$

et

$$D_C(\theta, t_0) = \sum_{n=0}^{\infty} p(n, qt_0) \frac{1}{n+1} \sum_{k=0}^n \frac{\partial}{\partial \theta} \tilde{P}^k(\theta) \quad (35)$$

Comme pour tout $k \in \mathbb{N}$,

$$\tilde{P}^k(\theta) = \tilde{P}^{k-1}(\theta) \tilde{P}(\theta), \quad \tilde{P}^0(\theta) = I$$

alors

$$\frac{\partial}{\partial \theta} \tilde{P}^k(\theta) = \left[\frac{\partial}{\partial \theta} \tilde{P}^{k-1}(\theta) \right] \tilde{P}(\theta) + \tilde{P}^{k-1}(\theta) \left[\frac{\partial}{\partial \theta} \tilde{P}(\theta) \right], \quad \frac{\partial}{\partial \theta} \tilde{P}^0(\theta) = 0$$

Désignons respectivement par $\bar{P}(\theta, t_0)$, $\bar{C}(\theta, t_0)$, $\bar{D}_P(\theta, t_0)$, et $\bar{D}_C(\theta, t_0)$ les matrices $P(\theta, t_0)$, $C(\theta, t_0)$, $D_P(\theta, t_0)$ et $D_C(\theta, t_0)$ calculées après troncature des sommes infinies au même rang N_0 tel que

$$\|E_P(0)\|_{\infty} = \|P(\theta, t_0) - \bar{P}(\theta, t_0)\|_{\infty} \leq 1 - \sum_{n=0}^{N_0} p(n, qt_0) \leq \varepsilon_0$$

La tolérance ε_0 est fixée à l'epsilon machine. Dans toute cette étude, les calculs sont effectués sur une station de travail SUN Ultra-30, 295 MHz, en double précision où $\varepsilon_0 \approx 10^{-16}$.

On peut démontrer facilement que la matrice d'erreur commise sur $C(\theta, t_0)$, notée $E_C(0)$, vérifie $\|E_C(0)\|_{\infty} \leq \varepsilon_0$. Les matrices d'erreur commise par la méthode SU sur le calcul de $D_P(\theta, t_0)$ et $D_C(\theta, t_0)$, notées respectivement $F_P(0)$ et $F_C(0)$, sont telles que :

$$\|F_P(0)\|_{\infty} \leq t_0 \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_{\infty} [\varepsilon_0 + p(N_0, qt_0)] \quad (36)$$

et

$$\|F_C(0)\|_{\infty} \leq \frac{t_0}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_{\infty} [\varepsilon_0 + p(N_0, qt_0)] \quad (37)$$

D'après les inégalités (28) et (37), l'erreur commise sur $S_Y(\theta, t_0)$, notée $F_Y(0)$, satisfait :

$$F_Y(0) \leq t_0 \|R\|_\infty \|F_C(0)\|_\infty$$

Le choix de $qt_0 < 1$ implique que $p(n, qt_0)$ décroît lorsque n croît. Le choix de $qt_0 = qt/2^m$ est celui qui minimise $N_0 + m$, c.à.d., la complexité temporelle de la méthode UP tout en assurant une précision acceptable. Ces critères sont vérifiés pour $qt_0 \leq 0.1$.

5.1 Majoration de l'erreur

Afin de pouvoir majorer l'erreur commise sur $D_C(\theta, t_l)$, $l = 1, \dots, m$, et déduire celle commise sur $S_Y(\theta, t_l)$, nous définissons d'abord les matrices (calculées) suivantes :

$$\begin{aligned} \bar{P}(\theta, l) &= \bar{P}(\theta, l-1)\bar{P}(\theta, l-1) \\ \bar{C}(\theta, l) &= \left[\frac{\bar{P}(\theta, l-1) + I}{2} \right] \bar{C}(\theta, l-1) \\ \bar{D}_P(\theta, l) &= \bar{D}_P(\theta, l-1)\bar{P}(\theta, l-1) + \bar{P}(\theta, l-1)\bar{D}_P(\theta, l-1) \\ \bar{D}_C(\theta, l) &= \frac{1}{2} [\bar{D}_P(\theta, l-1)] \bar{C}(\theta, l-1) + \frac{1}{2} [\bar{P}(\theta, l-1) + I] \bar{D}_C(\theta, l-1) \end{aligned}$$

La sensibilité calculée à l'instant t_l , notée $\bar{S}_Y(\theta, l)$, est telle que

$$\bar{S}_Y(\theta, l) = t\Pi(0) [\bar{D}_C(\theta, l)] R$$

Les matrices d'erreurs commises sur $P(\theta, t_l)$, $C(\theta, t_l)$, $D_P(\theta, t_l)$ et $D_C(\theta, t_l)$ sont respectivement notées $E_P(l)$, $E_C(l)$, $F_P(l)$ et $F_C(l)$.

On peut démontrer à l'ordre 1 que pour tout $l \geq 1$,

$$\|E_P(l)\|_\infty = \|P(\theta, t_l) - \bar{P}(\theta, l)\|_\infty \leq 2^{l+1}\varepsilon_0 \quad (38)$$

et

$$\|E_C(l)\|_\infty = \|C(\theta, t_l) - \bar{C}(\theta, l)\|_\infty \leq 2^{l+1}\varepsilon_0 \quad (39)$$

Il a été démontré [18] que

$$\|F_P(l)\|_\infty = \|D_P(\theta, t_l) - \bar{D}_P(\theta, l)\|_\infty \leq t_l \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty [2^l \varepsilon_0 + p(N_0, qt_0)] \quad (40)$$

Nous démontrons dans l'annexe A qu'à l'ordre 1

$$\begin{aligned} \|F_C(l)\|_\infty &= \|D_C(\theta, t_l) - \bar{D}_C(\theta, l)\|_\infty \\ &\leq \frac{t_l}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[\frac{1}{3}(2^{l+1} + \frac{1}{2^l})\varepsilon_0 + p(N_0, qt_0) \right] \end{aligned} \quad (41)$$

En prenant $l = m$, l'erreur $F_Y(m)$ commise sur $S_Y(\theta, t)$ vérifie alors

$$\begin{aligned} F_Y(m) &= S_Y(\theta, t) - \bar{S}_Y(\theta, m) \\ &\leq \frac{t^2}{2} \|R\|_\infty \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[\frac{1}{3}(2^{m+1} + \frac{1}{2^m})\varepsilon_0 + p(N_0, qt_0) \right] \end{aligned}$$

Signalons que si on remplace dans la relation (24) la tolérance ε par $\frac{1}{3}(2^{m+1} + \frac{1}{2^m})\varepsilon_0$ et $p(N_T, qt)$ par $p(N_0, qt_0)$, on retrouve le majorant de l'erreur donné par la méthode SU (cf. relation (24)).

5.2 Algorithme UP et complexité temporelle

L'algorithme présenté ci-dessous montre le calcul effectué par la méthode UP. Dans cet algorithme, les vecteurs $\bar{U}(\theta, l)$ et $\bar{V}(\theta, l)$ désignent respectivement $\bar{C}(\theta, t_l)R$ et $\bar{D}_C(\theta, t_l)R$, $l \geq 0$.

Entrée : t

Sortie : $\bar{S}_Y(\theta, t)$

Calculer $m = \min\{l/\frac{qt}{2^l} \leq 0.1\}$, $t_0 = t/2^m$

Calculer N_0 tel que $\|E_P(0)\|_\infty \leq \varepsilon_0$

Calculer $\bar{U}(\theta, 0)$, $\bar{V}(\theta, 0)$, $\bar{P}(\theta, 0)$, $\bar{D}_P(\theta, 0)$

Pour l depuis 1 jusqu'à m faire

$$\begin{aligned} \bar{V}(\theta, l) &= \frac{1}{2} [\bar{D}_P(\theta, l-1)] \bar{U}(\theta, l-1) + \frac{1}{2} [\bar{P}(\theta, l-1) + I] \bar{V}(\theta, l-1) \\ \bar{U}(\theta, l) &= \frac{1}{2} [\bar{P}(\theta, l-1) + I] \bar{U}(\theta, l-1) \end{aligned}$$

$$\begin{aligned}\bar{D}_P(\theta, l) &= \bar{D}_P(\theta, l-1)\bar{P}(\theta, l-1) + \bar{P}(\theta, l-1)\bar{D}_P(\theta, l-1) \\ \bar{P}(\theta, l) &= \bar{P}(\theta, l-1)^2\end{aligned}$$

Refaire

$$\bar{S}_Y(\theta, t) = t\Pi(0)\bar{V}(\theta, m)$$

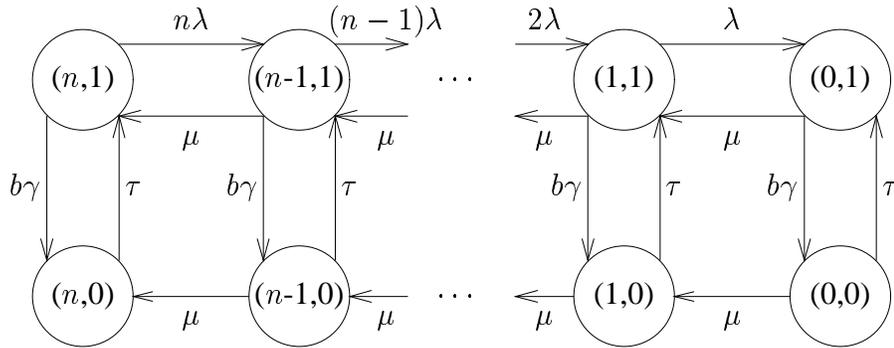
La méthode UP requiert essentiellement $3N_0$ produits vecteur-matrice et $3N_0$ produits matrice-matrice pour calculer $\bar{U}(\theta, 0)$, $\bar{V}(\theta, 0)$, $\bar{P}(\theta, 0)$ et $\bar{D}_P(\theta, 0)$. La complexité temporelle de cette phase est alors en $O(3N_0(M^3 + M^2))$, $N_0 < 10$. La phase des puissances, c.à.d., du retour à $\bar{V}(\theta, t)$ en partant de $\bar{U}(\theta, 0)$, $\bar{V}(\theta, 0)$, $\bar{P}(\theta, 0)$ et $\bar{D}_P(\theta, 0)$, requiert $3m$ produits vecteur-matrice et $3m$ produits matrice-matrice. La complexité temporelle de cette phase est alors en $O(3m(M^3 + M^2))$. La valeur de $\bar{S}_Y(\theta, t)$ se déduit par un simple produit scalaire. Malheureusement, les matrices $\bar{P}(\theta, l)$ et $\bar{D}_P(\theta, l)$, $l \geq 1$, se remplissent au fur et à mesure que l augmente. Par conséquent, le nombre des états doit être limité à quelques centaines d'états. D'après ce qui précède, la complexité temporelle de l'approche UP est donc en $O(3(N_0 + m)(M^3 + M^2))$.

6 Illustration numérique

6.1 Exemple

L'exemple considéré est un système informatique tolérant les panne constitué de n processeurs associés à b buffers (cases mémoires) [20]. Ce système peut être modélisé par une file d'attente $M/M/n/n + b$. Les arrivées ont lieu avec un taux Λ et sont perdues lorsqu'un buffer est plein; Θ dénote le taux de service.

Les processeurs (resp. les buffers) ont un taux de panne λ (resp. γ) et sont réparés individuellement avec un taux μ (resp. τ). Lorsqu'un processeur tombe en panne, le système continue de fonctionner et le nombre de processeurs diminue d'une unité. Le système est en panne lorsque tous les processeurs sont en panne ou dès que l'un des buffers tombe en panne. Le modèle est représenté par une CMTC dont le graphe de transition est donné sur la figure 1. L'espace d'état de la CMTC est $\mathbb{E} = \{(i, j); 0 \leq i \leq n, j = 0, 1\}$. La composante i de l'état (i, j) du système signifie que i processeurs sont opérationnels et la composante j est égal à zéro si au moins un des buffers est en panne, un sinon. Ainsi, l'ensemble des états opérationnels correspond à $j = 1$ pour tout $i = 0, \dots, n$.

FIG. 1 – Graphe de transition d'un système de n processeurs et b buffers

Le choix des taux de récompense des états dépend du type de mesure que l'on désire calculer. Celle-ci pouvant être une mesure de performance, de sûreté de fonctionnement ou de performabilité. Par exemple, pour calculer la disponibilité instantanée ou sur un intervalle du système, il suffit de choisir comme taux de récompense 1 pour les états opérationnels et 0 pour les états défectueux.

Pour notre exemple, nous associons à tous les états défectueux un taux de récompense égal à 0. À chaque état opérationnel, nous affectons un taux de récompense caractérisé par le nombre de processeurs en état de marche. Notons qu'il est également possible de considérer une capacité égale au nombre moyen de clients (tâches) dans le système à l'état stationnaire en considérant le nombre de processeurs en état de marche. Dans ce dernier cas, on trouve dans [21] une expression formelle de la capacité.

6.2 Résultats

Nous comparons ici le temps de calcul des trois méthodes présentées en fonction du nombre M des états et du temps de mission t du système considéré. Celui-ci possède $M = 2(n + 1)$ états et se trouve à l'instant $t = 0$ dans l'état $(n,1)$ avec une probabilité $\Pi_1(0) = 1$ ($\Pi_i(0) = 0, i \geq 2$). La valeur de $S_Y(\theta, t)$ est calculée pour $\theta = \lambda$. En prenant $\theta = \mu, \tau$ ou γ , les résultats obtenus vis-à-vis du temps de calcul sont similaires à ceux énoncés ci-dessous. Pour nos expérimentations, les valeurs des paramètres sont telles que $\lambda = \gamma = 10^{-6}$ et $\mu = \tau = 100$. Ceci conduit à un modèle markovien fortement raide. Ainsi par exemple, pour $t = 10^5$ et $M = 400$,

la valeur de qt est 10^8 environ. Les différents algorithmes sont exécutés sur une station de travail SUN Ultra-30, 295 MHz, en double précision où $\varepsilon_0 \approx 10^{-16}$. Pour la méthode SU, la tolérance donnée est 10^{-13} alors que la tolérance locale de IRK3 est égale à 10^{-10} . La figure 2 montre l'évolution de la sensibilité de l'ERC par rapport à λ en fonction de t . On constate que plus le temps de mission augmente, plus l'ERC est sensible à λ (baisse de l'ERC). La précision numérique des trois méthodes SU, IRK3 et UP a été traitée dans une étude précédente [10] et que cette précision est satisfaisante en pratique.

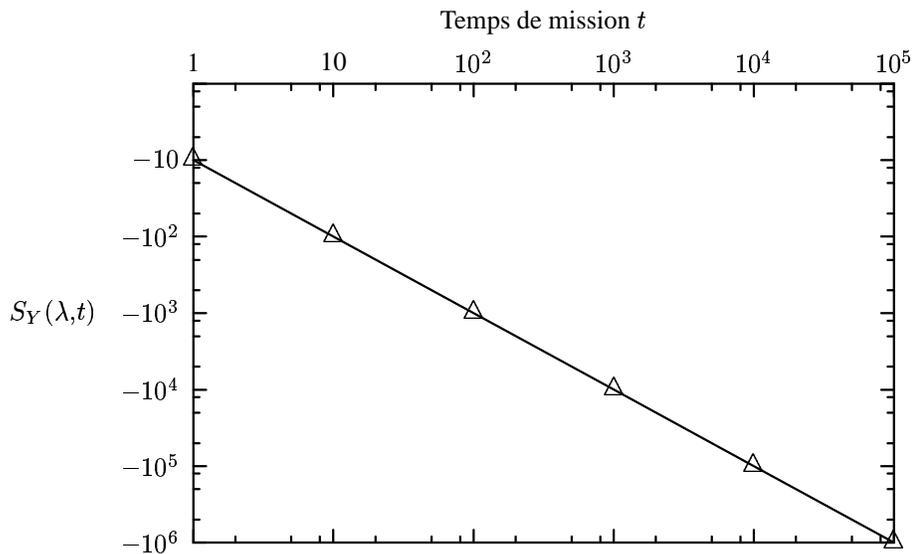


FIG. 2 – Sensibilité de l'ERC au taux de panne λ en fonction de t

Afin d'évaluer le temps de calcul des méthodes SU et IRK3, on effectue un stockage compact des matrices $Q(\theta)$ et de $\frac{\partial}{\partial \theta} Q(\theta)$. Les trois méthodes ont d'abord été exécutées pour $M = 10$ puis $M = 100$ en variant t de 1 à 10^5 . Pour $M = 10$, la méthode UP est la plus rapide. En prenant $M = 100$, la méthode SU reste la plus efficace pour $t \leq 100$, ses performances se dégradent lorsque t augmente. Cette méthode demande un temps de calcul prohibitif pour les grandes valeurs de t . Seules les méthodes UP et IRK3 sont retenues pour $M = 400$. La figure 5 montre que pour $t \leq 10^4$, la méthode IRK3 est plus rapide que la technique UP. Cette accélération est due à des valeurs du pas variant entre 10^{-4} et 10.87 en partant d'un pas initial

$h_0 = 10^{-7}$. Malgré cette accélération, IRK3 requiert 41 heures CPU pour $t = 10^5$ alors que UP demande près de 3 heures CPU. Il est donc recommandable d'utiliser IRK3 pour $t \leq 10^4$ et UP pour $t > 10^4$.

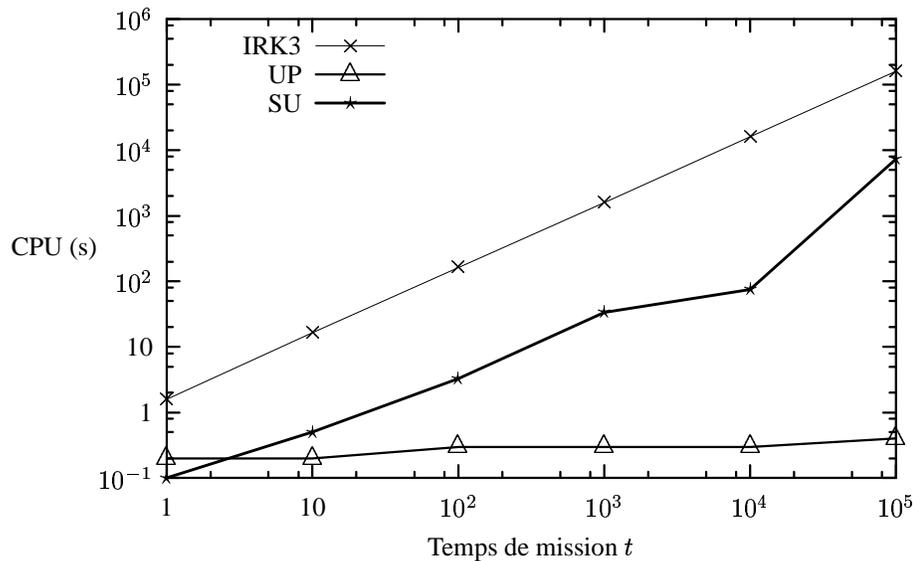
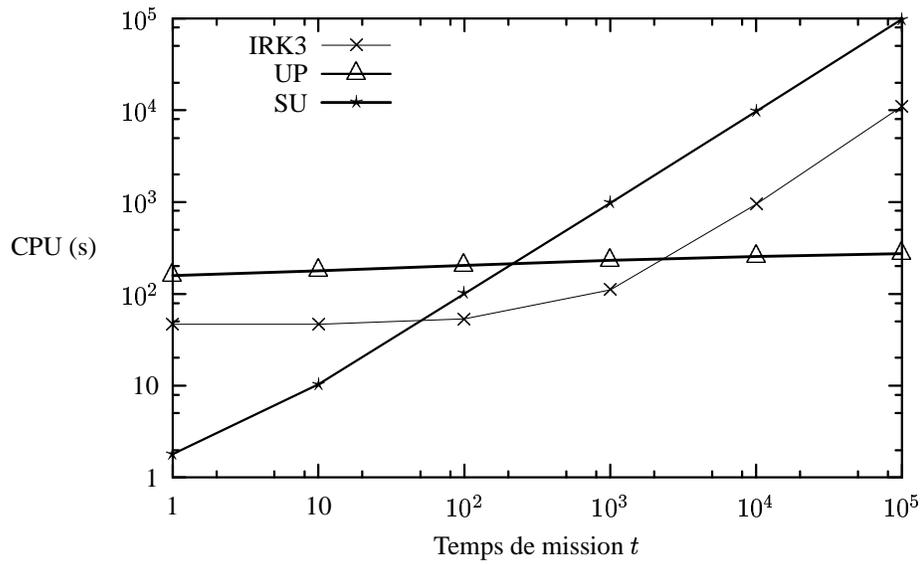
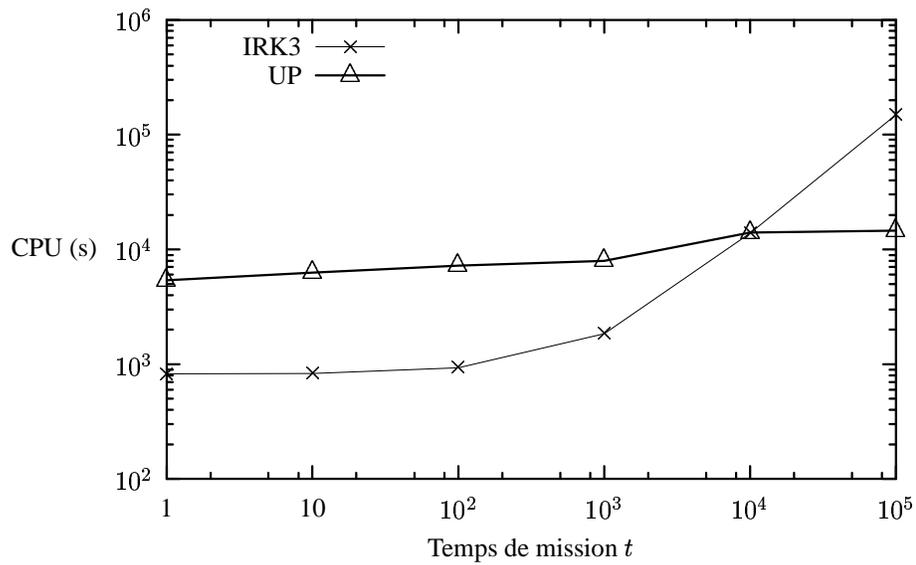


FIG. 3 – Temps CPU en fonction de t pour $M = 10$

7 Conclusion

Nous avons étudié dans ce rapport le calcul numérique de la sensibilité de l'ERC des modèles markoviens raides. Le principal problème auquel nous avons à faire face est le temps de calcul croissant avec la raideur. Après avoir étudié la méthode SU avec comme avantage le contrôle de l'erreur, nous avons considéré la méthode IRK3, une méthode ODE L-stable particulièrement adaptée aux modèles raides. Nous avons d'abord transformé le système différentiel donnant la sensibilité en un système ODE de la forme $y' = \lambda y$ où λ est une constante en vue d'appliquer la méthode IRK3. Nous avons ensuite amélioré cette dernière méthode avec un nouveau choix du pas garantissant l'erreur locale la plus proche de la tolérance locale

FIG. 4 – Temps CPU en fonction de t pour $M = 100$ FIG. 5 – Temps CPU en fonction de t pour $M = 400$

requis. Les valeurs du pas sont plus importantes que celles fournies par les algorithmes classiques.

Face aux difficultés relatives au contrôle de l'erreur globale par les méthodes ODE, nous avons proposé une approche basée sur la méthodologie UP. Cette méthode est particulièrement rapide pour les grandes valeurs du temps de mission t des systèmes à espace d'état raisonnable (≤ 400). Les 3 méthodes ont été comparées vis-à-vis du temps de calcul sur un exemple réel. Nous constatons que la méthode SU est intéressante pour les petites valeurs de t . Les méthodes UP et IRK3 sont recommandées pour les grandes valeurs de t .

Remerciements

Les auteurs tiennent à remercier Bernard Philippe, Directeur de Recherche à l'INRIA, et Raymond Marie, Professeur à l'université de Rennes 1, pour leurs précieuses remarques sur ce travail.

Références

- [1] J. T. Blake, A. L. Reibman and K. S. Trivedi. Sensitivity Analysis of Reliability and Performability Measures for Multiprocessor Systems. *Proceeding 1988 ACM SYGNETRICS, International Conference on Measurement and Modeling of Computer Systems, Santa Fe, New Mexico*, pages 177–186, 1988.
- [2] K. S. Trivedi and A. L. Riebman. Numerical Transient Analysis of Markov Models. *Computer and Operations Research*, 15(1):19–36, 1988.
- [3] P. W. Glynn. Stochastic Approximation for Monte-Carlo Optimization. *Proceedings of the 1986 Winter Simulation Conference, J. R. Wilson, J. O. Henriksen and S. D. Roberts (eds.) IEEE Press*, pages 356–365, 1986.
- [4] M. I. Reiman and A. Weiss. Sensitivity Analysis via Likelihood Ratios. *Proceedings of the 1986 Winter Simulation Conference, J. R. Wilson, J. O. Henriksen and S. D. Roberts (eds.) IEEE Press*, pages 285–289, 1986.
- [5] R. Marie, A. L. Reibman, and K. S. Trivedi. Transient analysis of acyclic markov chains. *Performance Evaluation*, 7(3):175–194, 1987.
- [6] M. Malhotra. A computationally efficient technique for transient analysis of repairable markovian systems. *Performance Evaluation*, 24(4):311–331, 1996.
- [7] C. Lindemann, M. Malhotra and K. S. Trivedi. Numerical Methods for Reliability Evaluation of Markov Closed Fault-Tolerant Systems. *IEEE Transactions on Reliability*, 44(4):694–704, 1995.
- [8] A. Reibman and K. Trivedi. Transient Analysis of Cumulative Measures of Markov Model Behavior. *Communication Statistics-Stochastic Models*, 5(4):683–710, 1989.
- [9] H. Abdallah and M. Hamza. Sensitivity analysis of instantaneous transient measures of highly reliable systems. 11th *European Simulation Symposium (ESS'99), Erlangen-Nuremberg, Germany, october 26-28*, pages 652–656, 1999.
- [10] H. Abdallah et M. Hamza. Sensibilité de mesures transitoires instantanées des systèmes informatiques hautement fiables. Publication Interne 1232, IRISA, Campus de Beaulieu, Rennes, France, Février 1999.

-
- [11] A. Reibman, R. Smith and K. Trivedi. Markov and Markov Reward Model Transient Analysis: An Overview of Numerical Approaches. *European Journal of Operational Research*, 40:257–267, 1989.
 - [12] P. Heidelberger and A. Goyal. Sensitivity Analysis of Continuous-time Markov Chains using Uniformization. *Computer Performance and Reliability*, editors: G. Iazeolla, P. J. Courtois and O. J. Boxma, Elsevier Science Publishers B. V. Amsterdam:93–104, 1988.
 - [13] R.S. Varga G. Birkhoff. Discretization errors for well set cauchy problems. *J. Math. Phys.*, 44:1–23, 1965.
 - [14] O. Axelsson. A class of a-stable methods. *BIT*, 9:185–199, 1969.
 - [15] S. P. Norsett E. Hairer and G. Wanner. *Solving ordinary differential equations*. Springer, New York, 1993.
 - [16] R. Bank, W. Courghan, W. Fichtner and E. Grosse, D. Rose, and R. Smith. Transient simulation of silicon devices and circuits. *IIE Transact. CAD*, 4:436–451, 1985.
 - [17] H. Abdallah and R. Marie. The Uniformized Power Method for Transient Solutions of Markov Processes. *Computers and Operations Research*, 20(5):515–526, April 1993.
 - [18] H. Abdallah. Sensitivity Computation of Reliability Markov Models Using the Uniformized Power Method. *Reliability Engineering and System Safety*, 56:53–59, 1997.
 - [19] H. Abdallah, R. Marie and B. Sericola. Point and Expected Interval Availability Analysis with Stationarity Detection. *Computers and Operations Research*, 26(3):211–218, October 1998.
 - [20] R. M. Smith V. G. Kulkarni, V. F. Nicola and K. S. Trivedi. Numerical evaluation of performability and job completion time in repairable fault-tolerant systems. *Proceedings IEEE 16-th Fault-Tolerant Computing Symposium, Vienna, Austria*, pages 252–257, July 1986.
 - [21] K. S. Trivedi. *Probability and Statistics with Reliability, Queueing and Computer Science Applications*. Prentice-Hall, Englewood Cliffs, N. J., 1982.

Annexe A

Majoration de l'erreur commise sur $D_C(\theta, t_l)$

Dans cette annexe, nous allons majorer à l'ordre 1 la la norme infinie de la matrice $F_C(l)$ d'erreur commise sur $D_C(\theta, t_l)$. On a :

$$\begin{aligned}\bar{D}_C(\theta, l) &= \frac{1}{2}[\bar{D}_P(\theta, l-1)\bar{C}(\theta, l-1) + \bar{P}(\theta, l-1)\bar{D}_C(\theta, l-1)] \\ &= \frac{1}{2}[(D_P(\theta, t_{l-1}) - F_P(l-1))(C(\theta, t_{l-1}) - F_C(l-1)) \\ &\quad + (P(\theta, t_{l-1}) - E_P(l-1) + I)(D_C(\theta, t_{l-1}) - F_C(l-1))] \\ &= D_C(\theta, t_l) - F_C(l)\end{aligned}$$

où

$$\begin{aligned}F_C(l) &= \frac{1}{2}[F_P(l-1)(C(\theta, t_{l-1}) - E_C(l-1)) + D_P(\theta, t_{l-1})E_C(l-1) \\ &\quad + (D(\theta, t_{l-1}) + I)F_C(l-1) + E_P(l-1)(D_C(\theta, t_{l-1}) - F_C(l-1))]\end{aligned}$$

A l'ordre 1, nous avons

$$\begin{aligned}\|F_C(l)\|_\infty &\leq \frac{1}{2}[\|F_P(l-1)\|_\infty \|C(\theta, t_{l-1})\|_\infty + \|D_P(\theta, t_{l-1})\|_\infty \|E_C(l-1)\|_\infty \\ &\quad + \|P(\theta, t_{l-1}) + I\|_\infty \|F_C(l-1)\|_\infty + \|E_P(l-1)\|_\infty \|D_C(\theta, t_{l-1})\|_\infty]\end{aligned}$$

Comme pour tout $l = 0, 1, \dots, m$, les matrices $C(\theta, t_l$ et $\frac{1}{2}(P(\theta, t_{l-1}) + I)$ sont stochastiques, il s'ensuit que :

$$\begin{aligned}\|F_C(l)\|_\infty &\leq \|F_C(l-1)\|_\infty \\ &\quad + \frac{1}{2}[\|F_P(l-1)\|_\infty + \|D_P(\theta, t_{l-1})\|_\infty \|E_C(l-1)\|_\infty \\ &\quad + \|E_P(l-1)\|_\infty \|D_C(\theta, t_{l-1})\|_\infty]\end{aligned}$$

D'après les inégalités (38), (39) et (40) et compte tenu du fait que

$$\|D_P(\theta, t_{l-1})\|_\infty \leq t_{l-1} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty$$

et

$$\|D_C(\theta, t_{l-1})\|_\infty \leq \frac{t_{l-1}}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty$$

nous avons

$$\|F_C(l)\|_\infty \leq \|F_C(l-1)\|_\infty + \frac{t_{l-1}}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty [2^{l+1}\varepsilon_0 + p(N_0, qt_0)]$$

avec

$$\|F_C(0)\|_\infty \leq \frac{t_0}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty [\varepsilon_0 + p(N_0, qt_0)] \quad (42)$$

Établissons maintenant par récurrence une majoration simple de $\|F_C(l)\|_\infty$. On a :

$$\|F_C(1)\|_\infty \leq \|F_C(0)\|_\infty + \frac{t_0}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty [2\varepsilon_0 + p(N_0, qt_0)]$$

et plus généralement,

$$\|F_C(l)\|_\infty \leq \|F_C(0)\|_\infty + \frac{1}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \sum_{k=0}^{l-1} t_k [2^{k+1}\varepsilon_0 + p(N_0, qt_0)]$$

Plus clairement,

$$\|F_C(l)\|_\infty \leq \|F_C(0)\|_\infty + \frac{1}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[\varepsilon_0 \sum_{k=0}^{l-1} 2^{k+1} t_k + p(N_0, qt_0) \sum_{k=0}^{l-1} t_k \right]$$

En remplaçant pour tout $k \geq 0$, t_k par $2^k t_0$, on obtient

$$\|F_C(l)\|_\infty \leq \|F_C(0)\|_\infty + \frac{1}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[\varepsilon_0 t_0 \sum_{k=0}^{l-1} 2^{2k+1} + p(N_0, qt_0) t_0 \sum_{k=0}^{l-1} 2^k \right]$$

ou encore

$$\|F_C(l)\|_\infty \leq \|F_C(0)\|_\infty + \frac{t_0}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[2\varepsilon_0 \frac{2^{2l} - 1}{3} + (2^l - 1)p(N_0, qt_0) \right]$$

En tenant compte de l'inégalité (42), nous avons

$$\|F_C(l)\|_\infty \leq \frac{t_0}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[\frac{(2^{2l+1} + 1)}{3} \varepsilon_0 + 2^l p(N_0, qt_0) \right]$$

ou plus simplement

$$\|F_C(l)\|_\infty \leq \frac{1}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty 2^l t_0 \left[\frac{1}{3} (2^{l+1} + \frac{1}{2^l}) \varepsilon_0 + p(N_0, qt_0) \right]$$

En remplaçant $2^l t_0$ par t_l , nous obtenons le résultat recherché

$$\|F_C(l)\|_\infty \leq \frac{t_l}{2} \left\| \frac{\partial}{\partial \theta} Q(\theta) \right\|_\infty \left[\frac{1}{3} (2^{l+1} + \frac{1}{2^l}) \varepsilon_0 + p(N_0, qt_0) \right]$$



Unité de recherche INRIA Rennes

IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur

INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-6399