Universiteit
Leiden
The Netherlands

De relatie tussen algoritmes voor het oplossen van Markovbeslissingsproblemen

Wel, W.J. van der

Citation
Wel, W. J. van der. (2005). *De relatie tussen algoritmes voor het oplossen van Markovbeslissingsproblemen*.

Version:	Not Applicable (or Unknown)
License:	License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository
Downloaded from:	https://hdl.handle.net/1887/3597566

Note: To cite this publication please use the final published version (if applicable).

DE RELATIE TUSSEN ALGORITMES VOOR HET OPLOSSEN VAN MARKOVBESLISSINGSPROBLEMEN

ARJAN VAN DER WEL

20 september 2005

Inhoudsopgave

Samenvatting

In deze scriptie worden een paar methoden besproken om Markovbeslissingsproblemen (MBP's) op te lossen, en van twee methoden wordt aangetoond dat ze equivalent zijn.

Hoofdstuk 1 begint met enkele modellen van MBP's, en de bijbehorende LP-formuleringen. Daarna worden twee inwendigpuntmethodes (IPM's) besproken waarmee LP-problemen opgelost kunnen worden.

Hoofdstuk 2 gaat over een nieuwe IPM voor het oplossen van MBP's. Dit is, tot zover bekend, het eerste 'streng polynomiale' algoritme voor het oplossen van MBP's.

Omdat gemiddelde optimaliteit niet altijd een bevredigend criterium is, wordt dit in hoofdstuk 3 uitgebreid met extra gevoelige criteria, waaronder n-verdisconteerde optimaliteit. Er wordt een algoritme besproken om m.b.v. strategieverbetering (SV) een n-verdisconteerd optimale strategie te vinden.

Hoofdstuk 4 gaat over een ander algoritme om een n-verdisconteerd optimale strategie te vinden: m.b.v. geneste LP-problemen (GLP), voor het irreduciblele geval.

In hoofdstuk 5 wordt aangetoond dat de algoritmes uit hoofdstuk 3 en 4 equivalent zijn. Tot nu toe was dat alleen voor gemiddelde optimaliteit bewezen.

Hoofdstuk 1

Inleiding

Dit hoofdstuk is gebaseerd op hoofdstuk 2 en 6 van [Kallenberg, 2001][1].

1.1 Markovbeslissingsketens

S is een eindige toestandsruimte.
In toestand i wordt een actie $a \in A(i)$ gekozen; er is dan een directe opbrengst $r_i(a)$, en een overgangskans $p_{ij}(a)$, $i, j \in S$.
$R = (\pi^1, \pi^2, \ldots)$ is een strategie; π^t is de beslisregel op tijdstip t;
π^t geeft de kans om een bepaalde actie te kiezen.
R is geheugenloos als π^t alleen van i_t (de toestand op tijdstip t) afhangt.
We schrijven in dit geval $\pi^t_{i_t a_t}$ voor de kans dat op tijdstip t in toestand i_t actie a_t wordt gekozen.
Als $\pi^t_{i_t a_t} \in \{0, 1\}$ $\forall a_t$, dan heet $\pi^t_{i_t a_t}$ deterministisch; we noteren zo'n beslisregel als een functie $f_t : S \to A$, waarbij $i \in S$ wordt afgebeeld op de actie $f_t(i)$ die met kans 1 wordt gekozen.
Een strategie met uitsluitend deterministische beslisregels heet een deterministische strategie. Een geheugenloze deterministische strategie noteren we als $R = (f_1, f_2, \ldots)$.
We kunnen bijna altijd met een geheugenloze deterministische strategie volstaan .
Als alle beslisregels identiek zijn, dan heet de strategie stationair. Zo'n strategie noteren we met f^∞.
Een geheugenloze stationaire gerandomiseerde (niet-deterministische) strategie $R = (\pi, \pi, \pi, \ldots)$ heeft beslisregel π die alleen afhangt van toestand i en actie a: $\pi : S \times A \to [0, 1]$. Zo'n strategie noteren we met π^∞.

1.1.1 Eindige horizon en totale opbrengsten

We beschouwen het systeem gedurende T perioden.
Bij start van het systeem op tijdstip 1 en strategie $R = (f_1, f_2, \ldots, f_T)$ is de totale verwachte opbrengst $v_i^T(R) = r_i(f_1) + [P(f_1)r(f_2)]_i + [P(f_1)P(f_2)r(f_3)]_i + \ldots + [P(f_1)P(f_2)\cdots P(f_{T-1}r(f_T)]_i,\ i \in S$.
$v_i^T := \max_R v_i^T(R),\ i \in S$, is de waardevector.
R is optimaal als $v_i^T(R) = v_i^T$ $\forall i \in S$.

1. L.C.M. Kallenberg, *Inleiding Besliskunde*, Univ. Leiden, 2001

Stelling 1.1 *Laat* $x_i^{T+1} = 0$, *en laat* f_t *en* x_t *voor* $t = T, T-1, \ldots, 2, 1$ *voldoen aan :*

$$x_i^t = r_i(f_t) + \sum_j p_{ij}(f_t) x_j^{t+1} = \max_{a \in A(i)} \{r_i(a) + \sum\nolimits_j p_{ij}(a) x_j^{t+1}\},\ i \in S.$$

Dan is $R_* = (f_1, f_2, \ldots, f_T)$ *een optimale strategie en* x^1 *is de waardevector.*

Bewijs
Met inductie naar T. ♢

1.1.2 Oneindige horizon en verdisconteerde opbrengsten

Per periode is het rentepercentage ρ.
Voor het waarderen in het heden van opbrengsten in de toekomst gebruiken we verdiscontering: een bedrag in periode t wordt vermenigvuldigd met $(1+\rho)^{-t}$ voor waardering in het heden. $\alpha := (1+\rho)^{-1}$ is de verdisconteringsfactor.
Met verdiscontering is de totale verdisconteerde opbrengst over een oneindige periode een eindig getal: als $|r_i(a)| \leq M\ \forall i \in S,\ a \in A(i)$, dan is de totale verdisconteerde opbrengst $\leq \frac{1}{1-\alpha} \cdot M$.
De verwachte verdisconteerde opbrengst $v_i^\alpha(R)$ is gedefinieerd als:

$$v_i^\alpha(R) := \sum_{t=1}^{\infty} \alpha^{t-1} \mathbf{E}_{i,R}[r_{X_t}(Y_t)],$$

met X_t en Y_t de stochastische varabelen voor de toestand resp. actie op tijdstip t.
Dus $v_i^\alpha(R)$ is ook te schrijven als:

$$v_i^\alpha(R) := \sum_{t=1}^{\infty} \alpha^{t-1} \sum\nolimits_{j,a} \mathbf{P}_{i,R}[X_t = j, Y_t = a] \cdot r_j(a).$$

$v_i^\alpha := \max_R v_i^\alpha(R)$ is de waardevector.
R_* heet een optimale strategie als $v_i^\alpha(R_*) = v_i^\alpha\ \forall i \in S$. Er kan bewezen worden dat er altijd een optimale, stationaire strategie bestaat.
De verwachte verdisconteerde opbrengst voor een stationaire strategie f^∞ met overgangsmatrix $P(f)$ en opbrengstvector $r(f)$ is (in vectornotatie):

$$v^\alpha(f^\infty) = [I - \alpha P(f)]^{-1} \cdot r(f) \tag{1.1}$$

Stelling 1.2 v^α *is de unieke oplossing van de optimaliteitsvergelijking*

$$x_i = \max_{a \in A(i)} \{r_i(a) + \alpha \sum\nolimits_j p_{ij}(a) x_j\},\ i \in S$$

Bewijs
M.b.v. de contraherende afbeelding $U : \mathbf{R}^N \to \mathbf{R}^N$:

$$(Ux)_i = \max_{a \in A(i)} \{r_i(a) + \alpha \sum_j p_{ij}(a) x_j\},\ i \in S$$

Een contraherende afbeelding heeft een uniek dekpunt x^*, en het blijkt dat $x^* = v^\alpha$. ♢

Een vector $v \in \mathbf{R}^N$ heet superharmonisch als

$$v_i \geq r_i(a) + \alpha \sum\nolimits_j p_{ij}(a) v_j \ \forall (i,a)$$

met $i \in S$ en $a \in A(i)$.

Stelling 1.3 *De waardevector v^α is de (componentsgewijs) kleinste superharmonische vector.*

Bewijs
v^α voldoet aan de optimaliteitsvergelijking en is dus superharmonisch. Als een vector v superharmonisch is, dan volgt uit de definitie en uit vergelijking (1.1) dat $v \geq v^\alpha(f^\infty)\ \forall f^\infty$, dus $v \geq v^\alpha$. ♢

Gevolg 1.1 *v^α is de unieke oplossing van het LP-probleem*

$$\min \left\{ \sum\nolimits_j \beta_j v_j \,\middle|\, \sum\nolimits_j [\delta_{ij} - \alpha p_{ij}(a)] v_j \geq r_i(a)\ \forall (i,a) \text{ met } i \in S \text{ en } a \in A(i) \right\} \quad (1.2)$$

met $\beta_j > 0\ \forall j \in S$. Het bijbehorende duale probleem is:

$$\max \left\{ \sum_{i,a} r_i(a) x_i(a) \,\middle|\, \begin{array}{l} \sum_{i,a} [\delta_{ij} - \alpha p_{ij}(a)] x_i(a) = \beta_j,\ j \in S \\ x_i(a) \geq 0\ \forall i \in S \text{ en } a \in A(i) \end{array} \right\} \quad (1.3)$$

Stelling 1.4 *Laat x^* een optimale oplossing zijn van (3); dan is de stationaire strategie f_*^∞ met $x_i^*(f_*(i)) > 0,\ i \in S$, goed gedefinieerd en een optimale strategie.*

Bewijs
M.b.v. de orthogonaliteitsrelaties. ♢

1.1.3 Oneindige horizon en gemiddelde opbrengsten (het irreducibele geval)

Definitie van de gemiddelde opbrengst :

$$\phi_i(R) := \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{E}_{i,R}[r_{X_t}(Y_t)] = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{j,a} \mathbf{P}_{i,R}[X_t = j, Y_t = a] \cdot r_j(a)$$

$\phi_i := \max_R \phi_i(R),\ i \in S$, is de waardevector.
R_* heet optimaal als $\phi_i(R_*) = \phi_i\ \forall i \in S$.
Er kan bewezen worden dat er altijd een optimale, stationaire strategie bestaat[2].

2. M.L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York, 1994, p. 450

Voor een stationaire strategie f^∞ is de gemiddelde opbrengst $\phi(f^\infty)$ in vectornotatie:

$$\phi(f^\infty) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} P^{t-1}(f) r(f) = P^*(f) r(f)$$

Er geldt het volgende verband tussen verdisconteerde en gemiddelde opbrengsten:

Stelling 1.5 *$\phi(f^\infty) = \lim_{\alpha\uparrow 1}(1-\alpha)v^\alpha(f^\infty)$ voor iedere stationaire strategie f^∞.*

Bewijs
Het bewijs is te vinden in het boek van Puterman. Stelling 8.2.3 op p. 341 zegt dat

$$v^\alpha(f^\infty) = \frac{1}{\alpha}\left[\frac{\alpha}{1-\alpha}\phi(f^\infty) + \sum_{n=0}^{\infty}\left(\frac{1-\alpha}{\alpha}\right)^n \cdot y_n\right]$$

voor $\rho = \frac{1-\alpha}{\alpha}$ klein genoeg en S een eindige toestandsruimte.
Hierbij is $y_n = (-1)^n H_P^{n+1} r,\ n = 0, 1, 2, \ldots$ en $H_P = (I - P + P^*)^{-1}(I - P^*)$.
Vermenigvuldig nu met $1-\alpha$ en laat α naar 1 stijgen; dit geeft:

$$\lim_{\alpha\uparrow 1}\phi(f^\infty) = \lim_{\alpha\uparrow 1}\left[(1-\alpha)v^\alpha(f^\infty) - (1-\alpha)\sum_{n=0}^{\infty}\left(\frac{1-\alpha}{\alpha}\right)^n \cdot y_n\right]$$

Hieruit volgt dat

$$\phi(f^\infty) = \lim_{\alpha\uparrow 1}(1-\alpha)v^\alpha(f^\infty)$$

♢

We nemen nu aan dat de overgangsmatrix $P(f)$ irreducibel is voor iedere stationaire strategie f^∞. Hieruit volgt dat de stationaire matrix voor iedere stationaire strategie f^∞ identieke rijen heeft, dus de vector $\phi(f)$ ook. Dus $\phi(f)$ kan als scalair beschouwd worden.

Bij dit probleem hoort de volgende LP-formulering:

$$\max\left\{\sum_{i,a} r_i(a)y_i(a) \;\middle|\; \begin{array}{lcll} \sum_{i,a}[\delta_{ij} - p_{ij}(a)]y_i(a) & = & 0, & j \in S \\ \sum_{i,a} y_i(a) & = & 1 & \\ y_i(a) \geq 0, & & & i \in S,\ a \in A(i) \end{array}\right\} \quad (1.4)$$

Het bijbehorende duale probleem is:

$$\min\left\{g \;\middle|\; \sum\nolimits_j [\delta_{ij} - p_{ij}(a)]h_j + g \geq r_i(a),\ i \in S,\ a \in A(i)\right\} \quad (1.5)$$

Stelling 1.6 *(i) (1.4) en (1.5) hebben optimale oplossingen y^* resp. (g^*, h^*), met $g^* = \phi$ en $\sum_a y_i^*(a) > 0\ \forall i \in S$;*
(ii) Als f_*^∞ een strategie is met $y_i^*(f_*(i)) > 0\ \forall i \in S$, dan is f_*^∞ een optimale

strategie;
(iii) De optimaliteitsvergelijking

$$g + h_i = \max_{a \in A(i)} \{r_i(a) + \sum\nolimits_j p_{ij}(a) h_j\},\ i \in S$$

heeft een oplossing waarin g uniek is en $g = \phi$. h is op een constante na bepaald.

Bewijsschets[3]
(i) Kies f^∞ willekeurig. $P(f)$ is irreducibel, dus $P^*(f)$ heeft identieke rijen; noteer deze rijvector met $\pi(f)$. Kies $y_i(a) = \begin{cases} \pi_i(f) & \text{als } a = f(i),\ i \in S; \\ 0 & \text{als } a \neq f(i),\ i \in S. \end{cases}$
$y_i(a)$ blijkt een toelaatbare oplossing van (1.4) te zijn.
Het toegelaten gebied is begrensd, dus (1.4) en (1.5) hebben een eindige optimale oplossing: y^* resp. (g^*, h^*). Laat y een toegelaten oplossing van (1.4) zijn, en $y_i = \sum_a y_i(a),\ i \in S$; laat π^∞ de gerandomiseerde stationaire strategie zijn met $\pi_i(a) = \begin{cases} \frac{y_i(a)}{y_i} & \text{als } y_i > 0,\ i \in S; \\ \text{willekeurig} & \text{als } y_i = 0,\ i \in S. \end{cases}$
Uit de eerste beperking van (1.4) volgt nu dat y een stationaire kansverdeling is mbt de Markovketen $P(\pi)$. $P(\pi)$ is irreducibel, dus $y_i > 0\ \forall i \in S$. Dit geldt ook voor de optimale oplossing $y^* \Rightarrow \sum_a y_i^*(a) > 0\ \forall i \in S$.
Uit de beperkingen van (1.5) volgt: $[I - P(f)]h^* + g^* e \geq r(f)$ voor iedere f^∞. Vermenigvuldigen met $P^*(f)$ geeft: $g^* e \geq P^*(f) r(f) = \phi(f^\infty) \Rightarrow g^* \geq \phi$.
Verder geldt dat er een strategie f_0^∞ (een verdisconteerd optimale strategie voor een verdisconteringsfactor voldoende dicht bij 1) bestaat zdd. $g^* \leq \phi(f_0^\infty)$. Er geldt namelijk dat

$$\phi(f_0) = \lim_{\alpha \uparrow 1} (1-\alpha) v^\alpha(f_0) \geq \liminf_{\alpha \uparrow 1} (1-\alpha) v^\alpha(R) \geq \phi(R)\ \forall R$$

De eerste gelijkheid is van Stelling 5.1; het bewijs van de eerste ongelijkheid is te vinden in het artikel van Blackwell[4]; de laatste ongelijkheid is de Tauberstelling (Stelling 1.7).
Dus $g^* = \phi$.
(ii) Uit de orthogonaliteitsrelaties volgt, na vermenigvuldiging met $P^*(f_*)$, dat $\phi \cdot e = P^*(f_*) r(f_*) \Rightarrow \phi = \phi(f_*)^\infty \Rightarrow f_*^\infty$ is optimaal.
(iii) Voor (g^*, h^*) geldt: $g^* + h_i^* \geq \max_{a \in A(i)} \{r_i(a) + \sum_j p_{ij}(a) h_j^*\},\ i \in S$.
Voor f_*^∞ geldt: $g^* + h_i^* = r_i(f_*) + \sum_j p_{ij}(f_*) h_j^*,\ i \in S$.
Dus $(g^*, h^*) = (\phi, h^*)$ is een oplossing van de optimaliteitsvergelijking. Stel dat (g, h) ook een oplossing is. Dan is $g \geq g^* = \phi$. Met een geschikte f^∞ volgt dat $ge = P^*(f) r(f) \Rightarrow g = \phi(f^\infty) \leq \phi$. Dus $g = \phi$.
Kies nu twee oplossingen van de optimaliteitsvergelijking: (ϕ, h^1) en (ϕ, h^2). Kies f zdd. $ge + h^1 = r(f) P(f) h^1$. Dan geldt dat $h^2 - h^1 \geq P(f)(h^2 - h^1)$, ofwel $(x := h^2 - h^1)$: $x - P(f)x \geq 0$. Ook is $P^*(f)[x - P(f)x] = 0 \Rightarrow x = P(f)x \Rightarrow x = P^*(f)x$. $P^*(f)$ heeft identieke rijen, dus $x = c \cdot e$, met c een constante. Dus $h^2 - h^1 = c \cdot e$.
$\diamondsuit$

3. Het bewijs staat in het dictaat van L.C.M. Kallenberg, *Inleiding Besliskunde*, Univ. Leiden, 2001
4. D. Blackwell, Discrete Dynamic programming, *Ann. Math. Stat.* **33**, 719-726 (1962)

Stelling 1.7 (Tauberstelling)[5]

$$\liminf_{\alpha\uparrow 1}(1-\alpha)v^{\alpha}(R) \geq \phi(R)\ \forall R$$

Bewijs
Kies een strategie R, een vaste begintoestand i en laat

$$x_t := \sum\nolimits_j \sum\nolimits_a \mathbf{P}_R[X_t = j, Y_t = a | X_1 = i] r_{ja},\ t \in \mathbf{N}.$$

Omdat de rij $\{X_t\}_{t=1}^{\infty}$ begrensd is, kunnen we voor $\alpha \in (0,1)$ schrijven:

$$(1-\alpha)^{-1} v_i^{\alpha}(R) = \left(\sum_{t=0}^{\infty} \alpha^t\right) \cdot \left(\sum_{t=1}^{\infty} \alpha^{t-1} x_t\right) = \sum_{t=1}^{\infty} \left(\sum_{s=1}^{t} x_s\right) \alpha^{t-1}$$

Ook geldt voor $\alpha \in [0,1)$ dat

$$(1-\alpha)^{-2} = \sum_{t=1}^{\infty} t\alpha^{t-1},$$

en dus ook $\phi_i(R) = (1-\alpha)^2 \phi_i(R) \sum_{t=1}^{\infty} t\alpha^{t-1}$. Voor $\alpha \in [0,1)$ kunnen we nu schrijven:

$$(1-\alpha)v_i^{\alpha}(R) - \phi_i(R) = (1-\alpha)^2 \sum_{t=1}^{\infty} \left\{ \frac{1}{t} \sum_{s=1}^{t} x_s - \phi_i(R) \right\} t\alpha^{t-1}$$

Kies nu een willekeurige positieve ϵ. Omdat $\phi_i(R) = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t$, is er een T_ϵ zdd. $\phi_i(R) < \frac{1}{T} \sum_{t=1}^{T} x_t + \frac{\epsilon}{2}\ \forall T > T_\epsilon$. We kunnen daarom schrijven:

$$(1-\alpha)^2 \sum_{t>T_\epsilon} \left\{ \frac{1}{t} \sum_{s=1}^{t} x_s - \phi_i(R) \right\} t\alpha^{t-1} > (1-\alpha)^2 \sum_{t>T_\epsilon} (-\frac{\epsilon}{2}) t\alpha^{t-1} \geq$$

$$-\frac{\epsilon}{2}(1-\alpha)^2 \sum_{t=1}^{\infty} t\alpha^{t-1} = -\frac{1}{2}\epsilon,$$

en

$$(1-\alpha)^2 \sum_{t=1}^{T_\epsilon} \left\{ \frac{1}{t} \sum_{s=1}^{t} x_s - \phi_i(R) \right\} t\alpha^{t-1} \geq (1-\alpha)^2 \min_{1\leq t\leq T_\epsilon} \left\{ \frac{1}{t} \sum_{s=1}^{t} x_s - \phi_i(R) \right\} \sum_{t=1}^{T_\epsilon} t\alpha^{t-1} > -\frac{1}{2}\epsilon$$

voor α voldoende dicht bij 1. Hiermee is bewezen dat

$$(1-\alpha)v_i^{\alpha}(R) - \phi_i(R) > -\epsilon$$

voor α voldoende dicht bij 1, dwz.

$$\liminf_{\alpha\uparrow 1}(1-\alpha)v^{\alpha}(R) \geq \phi(R).$$

$\diamondsuit$

5. Deze stelling staat in het collegedictaat van L.C.M. Kallenberg, *Stochastische Dynamische Programmering*, Univ. Leiden, 1982/1983

1.1.4 Optimaal stoppen van een Markovketen

Iedere toestand heeft nu twee acties: $A(i) = \{0, 1\},\ i \in S$. Als in toestand i actie 0 (stoppen) wordt gekozen, dan is er een directe opbrengst r_i en stopt het proces ($p_{ij}(0) = 0\ \forall j \in S$). Als actie 1 (doorgaan) wordt gekozen, dan is de opbrengst s_i en de overgangskansen zijn $p_{ij},\ j \in S$. Het doel is de totale opbrengst te maximaliseren. Dit model kan behandeld worden als een verdisconteerd model met $\alpha = 1$. Een bewijs hiervoor is te vinden in §7.2 van het boek van Puterman. Dit levert op:

$$\text{Optimaliteitsvergelijking: } v_i = \max\left\{r_i,\ s_i + \sum\nolimits_j p_{ij}v_j\right\},\ i \in S.$$

$$\text{LP-probleem: } \min\left\{\sum_j \beta_j v_j \;\middle|\; \begin{array}{rcl} v_i & \geq & r_i,\ i \in S \\ \sum_j[\delta_{ij} - p_{ij}]v_j & \geq & s_i,\ i \in S \end{array}\right\}.$$

$$\text{Duale probleem: } \max\left\{\sum_i r_i x_i + \sum_i s_i y_i \;\middle|\; \begin{array}{rcll} x_j + \sum_i[\delta_{ij} - p_{ij}]y_i & = & \beta_j, & j \in S \\ x_i, y_i & \geq & 0, & i \in S \end{array}\right\}.$$

1.2 Inwendigpuntmethodes

Inwendigpuntmethodes, kortweg IPM's, zijn methodes om LP-problemen op te lossen. In tegenstelling tot bij de simplexmethode wordt hierbij door het inwendige van het toegelaten gebied naar het optimum gelopen. We bekijken hier de affineschalingmethode en de centraalpadmethode.

1.2.1 De affieneschalingmethode

Bekijk het LP-probleem

$$\max\{p^T x | Ax = b;\ x \geq 0\} \tag{1.6}$$

De methode is iteratief; iedere oplossing x_k moet voldoen aan $x_k > 0$ en $Ax_k = b$, en $x^{k+1} := x^k + \lambda_k s^k$, met $\lambda_k \in \mathbf{R}^+$ de staplengte en $s^k \in \mathbf{R}^n$ de richtingsvector. Omdat $b = Ax^{k+1} = Ax^k + \lambda_k As^k = b + \lambda_k As^k$, is $As^k = 0$. Dus $s^k \in N(A) = \{s | As = 0\}$.
Als er nog geen toelaatbare inwendige oplossing x^0 bekend is, dan kunnen we het volgende doen:

1. Kies een $x^0 > 0$ en stel dat $Ax^0 \neq b$;
2. Definieer $y^0 := b - Ax^0$ en laat de scalair $z^0 = 1$;
3. Neem voor M een heel groot getal, start met (x^0, z^0) en los op:

$$\max\{p^T x - Mz | Ax + y^0 z = b;\ x, z \geq 0\}$$

(x^0, z^0) is een toelaatbaar inwendig punt, want $Ax^0 + y^0 z^0 = Ax^0 + (b - Ax^0) \cdot 1 = b$, en $x^0 > 0$ en $z^0 = 1 > 0$.
Er zijn twee mogelijkheden voor de optimale oplossing (x^*, z^*):

1. $z^* = 0:\ x^*$ is een optimale oplossing van het oorspronkelijke probleem;
2. $z^* > 0$: het oorspronkelijke probleem is ontoelaatbaar.

Bepaling van s
Er moet gelden dat $As = 0$, dus $s \in \mathcal{N}(A)$. Voor de projectiematrix $P = I - A^T(AA^T)^{-1}A$ geldt dat $Pw \in \mathcal{N}(A)\ \forall w \in \mathbf{R}^n$. Dus $s = Pw$ voldoet voor alle $w \in \mathbf{R}^n$.
Om P te kunnen berekenen, moet AA^T inverteerbaar zijn:

Lemma 1.1 *Als A een $m \times n$-matrix is met rang m, dan is AA^T inverteerbaar.*

Bewijs
M.b.v. lineaire algebra. ♢

We mogen aannemen dat het LP-probleem (1.6) de volledige rijrang heeft: $\text{rang}(A) = m$.

Het doel is nu om w zo te kiezen dat voor $s = Pw$ met $\|s\| = 1,\ p^T s$ maximaal is: bepaal

$$\max_s \{p^T s | \|s\|^2 = 1,\ As = 0\}$$

M.b.v. Lagrange-multiplicatoren is dit probleem om te zetten in:

$$\max_{s,\lambda,\mu} F(s,\lambda,\mu) = \max_{s,\lambda,\mu} \{p^T s - \lambda(s^T s - 1) - \mu^T As\} \tag{1.7}$$

Afgeleiden naar s, λ en μ nulstellen geeft:

$$p - 2\lambda s - A^T\mu = 0;\ s^T s = 1;\ As = 0. \tag{1.8}$$

$\|s\|$ is eigenlijk niet van belang, dus we mogen λ willekeurig kiezen. Neem $\lambda = \frac{1}{2}$; dit geeft: $p - s - A^T\mu = 0 \Rightarrow Ap - AA^T\mu = 0 \Rightarrow \mu = (AA^T)^{-1}Ap$. Dus

$$s = p - A^T\mu = Pp \tag{1.9}$$

Bepaling λ_k
Er moet gelden dat $x^{k+1} = x^k + \lambda_k s^k > 0$, dus

$$\lambda_k < \min_j \left\{ -\frac{x_j^k}{s_j^k} \,\middle|\, s_j^k < 0 \right\} =: \rho_k.$$

Dus als $s^k \geq 0$, dan is er een oneindige oplossing.
Kies $\lambda_k = \alpha_k \rho_k$ met $\alpha_k \in (0,1)$. Vaak wordt α dicht bij 1 gekozen, bijv. $\alpha = 0.95$.

Schaling
Aan de rand van het assenstelsel is de stapgrootte meestal klein. Een oplossing is: schaal de variabelen zdd. x^k zo ver mogelijk van alle assen af komt te liggen, bereken de beste nieuwe oplossing en transformeer terug.
Definieer daarom $\xi_j = \frac{x_j}{x_j^k},\ j = 1, 2, \ldots, n$. Dan is $\xi_j^k = 1\ \forall j$, dus ξ_j^k ligt even ver van alle assen. Definieer verder de diagonaalmatrix D door $d_{ii} = x_i^k,\ i = 1, 2, \ldots, n$. Dan geldt dat $x = D\xi$, en $x \geq 0 \Leftrightarrow D\xi \geq 0 \Leftrightarrow \xi \geq 0$.
Het LP-probleem wordt nu:

$$\max\{(Dp)^T \xi | AD\xi = b,\ \xi \geq 0\} \tag{1.10}$$

De richtingsvector wordt:

$$\sigma = Dp - DA^T[AD^2A^T]^{-1}AD^2p$$

Transformatie terug:

$$s = D\sigma = D^2[p - A^Tu] \text{ met } u = [AD^2A^T]^{-1}AD^2p$$

Voor de beste keuze voor s moeten we oplossen:

$$\max\{p^Ts | As = 0;\ \|D^{-1}s\| \leq 1\} \tag{1.11}$$

Stelling 1.8 *De optimale oplossing van (1.11) is* $s^* = \frac{D^2(p-A^Tu)}{\|D(p-A^Tu)\|}$, *met* $u = (AD^2A^T)^{-1}AD^2p$.

Bewijs
M.b.v. de ongelijkheid van Schwarz wordt aangetoond dat $p^Ts \leq \|D(p - A^Tu)\|$ voor een willekeurige toelaatbare richtingsvector s.
Anderzijds is s^* toelaatbaar en $p^Ts^* = \|D(p - A^Tu)\|$ ♢

Opmerking: we delen door $\|D(p - A^Tu)\|$, dus moet gelden dat $p \neq A^Tu$. Maar als $p = A^Tu$, dan is de huidige oplossing al optimaal: de richtingsvector is dan bij schaling gelijk aan $s = D\sigma = D^2(p - A^Tu) = 0$.

Convergentie

Stelling 1.9 *Laat* $x^{k+1} = x^k + \lambda\frac{D_k^2(p-A^Tu^k)}{\|D^k(p-A^Tu^k)\|}$, *met* $u^k = (AD_k^2A^T)^{-1}AD_k^2p$. *Dan geldt:*
(i) $0 < \lambda \leq \frac{2}{3}$ *(kleinestapversie)* $\Rightarrow$ x^k *en* u^k *convergeren naar de optimale oplossingen van LP resp. DLP;*
(ii) Als $\lambda > \frac{2}{3}$ *en iedere basisoplossing is primaal en duaal niet-gedegenereerd, dan convergeren* x^k *en* u^k *naar de optimale oplossingen van LP resp. DLP;*
(iii) Er is een voorbeeld met $\lambda > \frac{2}{3}$ *waarin de methode convergeert naar een niet-optimale oplossing.*

Bewijs
Het bewijs is te vinden in de secties 9.1 en 9.2 van het boek van Bertsimas en Tsitsiklis[6]. ♢

Opmerkingen
1. Na zekere tijd ligt x^k in de buurt van een hoekpunt.
2. Ieder niet-optimaal hoekpunt is afwijzend.
3. Ieder optimaal hoekpunt is aantrekkend.
4. Om praktische redenen wordt λ meestal dicht bij 1 gekozen. Voorbeelden waarin geen convergentie optreedt, komen in de praktijk (bijna) niet voor.

Stelling 1.10 *In de buurt van een hoekpunt is* u *de vector van de duale variabelen.*

6. D. Bertsimas en J.N. Tsitsiklis, *Introduction to linear programming*, Athena Scientific, 1997

Bewijs
We kunnen de matrix A opsplitsen: $A = (B, N)$ met B de basismatrix van het hoekpunt. Dan is $D^2 = \begin{pmatrix} D_B^2 & \\ & 0 \end{pmatrix}$, $AD^2A^T = BD_B^2B^T$, $AD^2p = [AD^2A^T][B^T]^{-1}p_B$ en $u = [B^T]^{-1}p_B$.
Het duale stelsel $A^Tu - v = p$ is te schrijven als $\begin{pmatrix} B^T \\ N^T \end{pmatrix} u - \begin{pmatrix} v_B \\ v_N \end{pmatrix} = \begin{pmatrix} p_B \\ p_N \end{pmatrix}$.
Hieruit volgt dat $B^Tu - v_B = p_B$. $x_B > 0$, dus op grond van de orthogonaliteitsrelatie $v^Tx = 0$ is $v_B = 0$. Dus $B^Tu = p_B \Rightarrow u = [B^T]^{-1}p_B$. ♢

De vector $u = (AD^2A^T)^{-1}AD^2p$ kan worden gevonden als unieke oplossing van het stelsel $[AD^2A^T]u = AD^2p$. De matrix AD^2A^T is symmetrisch. Voor het oplossen van zo'n symmetrisch stelsel zijn speciale numerieke technieken beschikbaar.
Omdat $p - A^Tu = v$, wordt $s = -D^2v$.

Stopcriterium
p^Tx^k en b^Tu^k convergeren naar elkaar. Kijk naar het relatieve verschil: stop als

$$\frac{|b^Tu^k - p^Tx^k|}{1 + |p^Tu^k|} \leq \epsilon.$$

We krijgen zo het volgende algoritme:

1. (a) Start met $x^0 > 0$ en $Ax^0 = b$, waarbij A de volledige rijrang heeft.
   (b) $k = 0$ en kies $\epsilon > 0$.
2. (a) Bepaal de diagonaalmatrix D met $d_{ii} = x_i^k,\ i = 1, \ldots, n$.
   (b) Bepaal u^k door het stelsel $[AD^2A^T]u = AD^2p$ op te lossen.
   (c) $v^k = A^Tu^k - p$ en $s^k = -D^2v^k$.
   (d) Als $s^k \geq 0$, dan is er een oneindige oplossing: STOP.
   Anders: $\rho_k = \min\left\{ \frac{-x_j^k}{s_j^k} \,\middle|\, s_j^k < 0 \right\}$, kies $\alpha_k \in (0,1)$, laat $\lambda_k = \alpha_k\rho_k$ en $x^{k+1} = x^k + \lambda_k s^k$.
3. Als $\frac{|b^Tu^k - p^Tx^k|}{1+|p^Tx^k|} \leq \epsilon$: stop (neem x^{k+1} resp. u^k als oplossingen voor LP resp. DLP).
   Anders: $k := k + 1$ en ga naar stap 2.

1.2.2 De centraalpadmethode

Beschouw

$$\text{LP: } \max\{p^Tx | Ax = b;\ x \geq 0\}$$

en

$$\text{DLP: } \min\{b^Tu | A^Tu - v = p;\ v \geq 0\},$$

met A een $m \times n$-matrix met rang m. Ieder LP-probleem is zo te formuleren. Bij de centraalpadmethode laten we de beperkingen $x_j \geq 0,\ 1 \leq j \leq n$, weg. Om $x_j,\ 1 \leq j \leq n$, toch positief te houden, breiden we de doelfunctie uit met de term $\mu \sum_{j=1}^n \log x_j,\ \mu > 0$. Deze uitgebreide doelfunctie noemen we de logaritmische barrièrefunctie:

$$B_\mu(x) = p^Tx + \mu \sum_{j=1}^n \log x_j \tag{1.12}$$

Zo ontstaat het niet-lineaire programmeringsprobleem:

$$\max\{B_\mu(x)|Ax = b\} \tag{1.13}$$

De optimale oplossing van (1.13) is ongeveer gelijk aan de optimale oplossing van het oorspronkelijke probleem als μ dicht bij 0 ligt.
Voor de doelfunctie $B_\mu(x)$ geldt:
$\nabla^2 B_\mu(x) = -\mu \cdot X^{-2};\ \mu > 0$ en $[X^{-2}]_{ij} = \begin{cases} x_i^2 > 0 & \text{als } i = j \\ 0 & \text{anders} \end{cases}$
Dus X^{-2} is een diagonaalmatrix met positieve diagonaalelementen. Dus X^{-2} is positief definiet. Dus $\nabla^2 B_\mu(x) = -\mu \cdot X^{-2}$ is negatief definiet. Dus $B_\mu(x) = p^T x + \mu \sum_{j=1}^n \log x_j$ is strict concaaf.
Dus (1.13) heeft een unieke optimale oplossing die wordt bepaald door de KKT-voorwaarden:

Definitie KKT-voorwaarden
Voor het niet-lineaire optimaliseringsprobleem

$$\max\left\{ f(x) \;\middle|\; \begin{array}{ll} g_i(x) \le 0, & 1 \le i \le p \\ g_i(x) = 0, & p+1 \le i \le m \end{array} \right\}, \tag{1.14}$$

met f en g_i minstens één keer continu differentieerbaar, en x^* een lokaal optimum van (1.14), zijn er vectoren u^* en v^* zdd.:

1. $\nabla f(x^*) = \sum_{i=1}^p u_i^* \nabla g_i(x^*) + \sum_{i=p+1}^m v_i^* \nabla g_i(x^*)$;
2. $u_i^* g_i(x^*) = 0,\ 1 \le i \le p$;
3. $u_i^* \ge 0,\ 1 \le i \le p$.

Bij (1.13) zijn de KKT-voorwaarden, inclusief toelaatbaarheid: $\exists u \in \mathbf{R}^m$ zdd. $Ax = b; p_j + \frac{\mu}{x_j} = \sum_{i=1}^n u_i a_{ij},\ 1 \le j \le n$. In matrixnotatie:

$$Ax = b,\ A^T u - \mu \cdot X^{-1} e = p, \tag{1.15}$$

met X een diagonaalmatrix met $x_j,\ 1 \le j \le n$, op de diagonaal, en $e = (1, 1, \ldots, 1)^T$.

Het duale probleem $\min\{b^T u | A^T u - v = p,\ v \ge 0\}$ is analoog te benaderen met het strict convexe probleem

$$\min\left\{\ b^T u - \mu \cdot \textstyle\sum_{j=1}^n \log v_j \ \middle|\ A^T u - v = p\ \right\} \tag{1.16}$$

De KKT-voorwaarden zijn hier: $\exists x \in \mathbf{R}^n$ zdd. $A^T u - v = p;\ b_i = \sum_{j=1}^n x_j a_{ij},\ 1 \le i \le m;\ -\frac{\mu}{v_i} = \sum_{j=1}^n x_j(-\delta_{ij}) = -x_i,\ 1 \le i \le n$. In matrixnotatie:

$$A^T u - v = p;\ Ax = b;\ XVe = \mu \cdot e, \tag{1.17}$$

met V een diagonaalmatrix met $v_j,\ 1 \le j \le n$, op de diagonaal.
$v = Ve = \mu \cdot X^{-1} e$, dus geldt: x is optimaal voor (1.13) en (u, v) voor (1.16) $\Leftrightarrow$ (1.17) geldt.

IPM om (1.13) op te lossen
$B_\mu(x)$ heeft gradiënt $\nabla B_\mu(x) = p + \mu \cdot X^{-1} e$ en Hessiaan $\nabla^2 B_\mu(x) = -\mu \cdot X^{-2}$. We kunnen $B_\mu(x + s)$ met een tweede orde Taylorbenadering benaderen:

$$B_\mu(x + s) \approx B_\mu(x) + (p + \mu \cdot X^{-1} e)^T s - \frac{1}{2} \mu \cdot s^T X^{-2} s$$

Er moet gelden dat $As = 0$. De beste richting is de vector s die $B_\mu(x+s)$ maximaliseert, dus de oplossing van:

$$\max\left\{\ (p+\mu\cdot X^{-1}e)^T s - \tfrac{1}{2}\mu\cdot s^T X^{-2}s \ \middle|\ As=0\ \right\} \tag{1.18}$$

Voor de doelfunctie $f(s)$ van (1.18) geldt:
$\frac{\partial^2 f(s)}{\partial s_i^2} = -\frac{\partial^2}{\partial s_i^2}(-\frac{1}{2}\mu\cdot s^T X^{-2}s) = -\frac{1}{2}\mu\cdot\frac{\partial^2}{\partial s_i^2}\left(\frac{s_i^2}{x_i^2}\right) = -\frac{1}{2}\mu\cdot\frac{2}{x_i^2} = -\frac{\mu}{x_i^2}$. Dus $\nabla^2 f(s) = -\mu X^{-2}$ en deze matrix is negatief. Dus $f(s)$ is strict concaaf.
Dus de KKT-voorwaarden zijn nodig en voldoende voor optimaliteit:

$$\mu\cdot X^{-1}s + A^T u = p + \mu\cdot X^{-1}e;\ As = 0 \tag{1.19}$$

Dit stelsel heeft $n+m$ onbekenden, namelijk s en u. Het is als volgt analytisch op te lossen:
Vermenigvuldig de eerste vergelijkingen met AX^2 en gebruik dat $As = 0$. Dit geeft

$$u = (AX^2A^T)^{-1}A(X^2p + \mu\cdot Xe) \tag{1.20}$$

Vermenigvuldig de eerste vergelijkingen nu met X^2, gebruik (1.20) en $v = A^T u - p$. Dit geeft

$$s = \frac{-X^2 v + \mu\cdot Xe}{\mu} \tag{1.21}$$

Iteratiestap
Begin met een $x^k > 0$ met $AX^k = b,\ v^k > 0$ en u^k zdd. $A^T u^k - v^k = p,\ \mu_k > 0,\ \alpha\in(0,1)$ en $\epsilon > 0$:

1. Optimaliteitstest:
   Als $(v^k)^T x^k \le \epsilon$: stop (x^k en (u^k, v^k) zijn de (benaderende) oplossingen van het LP- resp. DLP-probleem).
   Anders: ga naar stap 2.
2. (a) $\mu_{k+1} = \alpha\cdot\mu_k$.
   (b) $u^{k+1} = (AX_k^2A^T)^{-1}A[X_k^2 p + \mu_{k+1}\cdot X_k e];\ v^{k+1} = A^T u^{k+1} - p;\ s^k = \frac{-X_k^2 v^{k+1} + \mu_{k+1}\cdot X_k e}{\mu_{k+1}}$.
   (c) $x^{k+1} = x^k + s^k$.
   (d) $k := k+1$ en ga naar stap 1.

Keuze van α
In theorie blijkt de keuze

$$\alpha = 1 - \frac{\sqrt{\beta}-\beta}{\sqrt{\beta}+\sqrt{n}}, \tag{1.22}$$

met $\beta\in(0,1)$ willekeurig, goed te werken:

Stelling 1.11 *Veronderstel dat we starten met $x^0 > 0$ zdd. $Ax^0 = b$, een $v^0 > 0$ en u^0 zdd. $A^T u^0 - v^0 = p$, een $\beta\in(0,1)$ en een $\alpha = 1 - \frac{\sqrt{\beta}-\beta}{\sqrt{\beta}+\sqrt{n}}$, en een $\mu_0 > 0$ zdd. $\|\frac{1}{\mu_0}X_0V_0e - e\| \le \beta$. Dan stopt het algoritme na maximaal K iteraties met primaal en duaal toelaatbare oplosingen x^K en (u^K, v^K) met $\sum_{j=1}^n x_j v_j \le \epsilon$ en $K = \left\lceil \frac{1}{1-\alpha}\ln\left\{\frac{1+\beta}{1-\beta}\cdot\frac{(x^0)^T v^0}{\epsilon}\right\}\right\rceil$.*

Bewijsschets
M.b.v. inductie naar k wordt eerst aangetoond dat x^k en (u^k, v^k) toelaatbare oplossingen van het LP- resp. DLP-probleem zijn met $x^k > 0,\ v^k > 0$ en $\|\frac{1}{\mu_k} X_k V_k e - e\| \leq \beta\ \forall k$.
Uit $\|\frac{1}{\mu_k} X_k V_k e - e\| \leq \beta\ \forall k$ is af te leiden dat $n\mu_k(1-\beta) \leq (x^k)^T v^k \leq n\mu_k(1+\beta)$.
M.b.v. $1 - x \leq e^{-x}\ \forall x \geq 0$ is verder af te leiden dat $\mu_k \leq e^{-k(1-\alpha)} \cdot \mu_0$.
Hieruit volgt dat $(x^k)^T v^k \leq \epsilon$ als $k \geq \frac{1}{1-\alpha} \ln \frac{n(1+\beta)\mu_0}{\epsilon}$.
$n(1-\beta)\mu_0 \leq (x^0)^T v^0$, dus het algoritme stopt als $k \geq \left\lceil \frac{1}{1-\alpha} \ln \left\{ \frac{1+\beta}{1-\beta} \cdot \frac{(x^0)^T v^0}{\epsilon} \right\} \right\rceil$. $\diamondsuit$

Het vinden van een oplossing die aan de voorwaarden van Stelling 1.11 voldoet
We nemen aan dat alle getallen in $A,\ p$ en b geheeltallig zijn en in absolute waarde begrensd door U. Er kan bewezen worden dat dan iedere component van een basisoplossing x begrensd wordt door $(mU)^m$. Dus $e^T x \leq n(mU)^m$.
Dus het oorspronkelijke LP-probleem is equivalent met

$$\max\{p^T x \mid Ax = b;\ e^T x \leq n(mU)^m;\ x \geq 0\} \tag{1.23}$$

Met de notaties $\overline{b} = c \cdot b$ en $c = \frac{n+2}{n(mU)^m}$ is (1.23) equivalent met

$$\max\{p^T x \mid Ax = \overline{b};\ e^T x \leq n+2;\ x \geq 0\} \tag{1.24}$$

Laat M een zeer groot getal zijn, en beschouw het probleem

$$\max\{p^T x - M x_{n+1} | Ax + (\overline{b} - Ae)x_{n+1} = \overline{b};\ e^T x + x_{n+1} + x_{n+2} = n+2;\ x_1, \ldots, x_{n+2} \geq 0\} \tag{1.25}$$

Het bijbehorende duale probleem is

$$\min \left\{ \overline{b}^T u + (n+2)u_{m+1} \;\middle|\; \begin{array}{rcrcrcl} A^T u & + & u_{m+1} e & - & v & = & p \\ (\overline{b} - Ae)^T u & + & u_{m+1} & - & v_{n+1} & = & -M;\ v_1, \ldots, v_{n+2} \geq 0 \\ & & u_{m+1} & - & v_{n+2} & = & 0 \end{array} \right\} \tag{1.26}$$

Neem nu $\mu_0 = \frac{1}{\beta}\sqrt{\|p\|^2 + M^2}$ en $x_j^0 = 1$ voor $1 \leq j \leq n+2$. Dan is x^0 een strict positieve toelaatbare oplossing voor (1.25) met bijbehorende $X_0 = I$. Neem

$$u_i^0 = \begin{cases} 0 & \text{voor} \quad 1 \leq i \leq m \\ \mu_0 & \text{voor} \quad i = m+1 \end{cases}$$

en

$$v_j^0 = \begin{cases} -p_j + \mu_0 & \text{voor} \quad 1 \leq j \leq n \\ M + \mu_0 & \text{voor} \quad j = n+1 \\ \mu_0 & \text{voor} \quad j = n+2 \end{cases};$$

dan is (u^0, v^0) toelaatbaar voor (1.26) en v is strict positief. Verder geldt:

$$\left\| \frac{1}{\mu_0} X_0 V_0 e - e \right\| = \left\| \frac{1}{\mu_0} V_0 e - e \right\| = \frac{1}{\mu_0} \|(V_0 - \mu_0 I)e\| = \frac{1}{\mu_0}\sqrt{\|p\|^2 + M^2} = \beta.$$

De bovenstaande startwaarden voldoen dus aan de voorwaarden van Stelling 4.13. Omdat M zeer groot is, is in de optimale oplossing van (1.25), $x_{n+1} = 0$, en deze oplossing is dus ook optimaal voor het oorspronkelijke probleem.

Hoofdstuk 2

Een nieuwe IPM voor het oplossen van MBP's

2.1 Inleiding

We gebruiken hier de volgende LP-formulering:

$$\text{Primaal: } \min\{c^T x \mid Ax = b,\ x \geq 0\} \tag{2.1}$$

$$\text{Duaal: } \max\{b^T y \mid A^T y + s = c,\ s \geq 0\} \tag{2.2}$$

waarbij $A \in \mathbf{R}^{m \times n}$, en A heeft rang m; $c \in \mathbf{R}^n,\ b \in \mathbf{R}^m$ zijn gegeven vectoren, en $x \in \mathbf{R}^n,\ y \in \mathbf{R}^m,\ s \in \mathbf{R}^n$ zijn onbekende vectoren; s is de duale verschilvariabele. We noteren dit probleem met $LP(A, b, c)$.

Het Vavasis-Ye-algoritme[1] is een inwendigpuntalgoritme dat kleine stappen afwisselt met grotere layered least squares (LLS)-stappen om het centrale pad te volgen. Dit algoritme noemen we layered-step interior point, LIP. Het eindigt na een eindig aantal stappen, en het aantal iteraties hangt alleen van A af. Bovendien is het aantal iteraties polynomiaal in n. Er zijn ook IPM's waarbij de complexiteit ook van de vectoren b en c afhangt. Dit is van belang, want bij veel problemen gedraagt A zich goed, maar b en c zijn willekeurige vectoren.

2.2 Het Markovbeslissingsprobleem

Het Markovbeslissingsprobleem (MBP) kan als volgt geformuleerd worden:

$$\min \left\{ \sum_{j=1}^{n} c_j^T x_j \;\middle|\; \begin{array}{l} \sum_{j=1}^{n} (E_j - \alpha P_j) x_j = e \\ x_j \geq 0,\ 1 \leq j \leq n \end{array} \right\},$$

waarbij E_j de $(n \times k)$-matrix is met op de j-de rij allemaal enen en verder nullen, en P_j een $(n \times k)$-matrix zdd. $e^T P_j = e^T$ en $P_j \geq 0\ \forall j$; $x_j \in \mathbf{R}^k$ correspondeert met de beslissingsvariabelen die horen bij de k acties bij toestand j, en c_j is de bijbehorende kostenvector. Het bijbehorende duale probleem is

$$\max\{e^T y \mid (E_j - \alpha P_j)^T y \leq c_j,\ 1 \leq j \leq n\}$$

1. S. Vavasis en Y. Ye, A primal-dual interior-point method whose running time depends only on the constraint matrix, *Mathematical Programming* **74**, p. 79-120, 1996

Door de beslissingsvariabelen te sorteren naar acties, krijgen we de volgende formulering:

$$\text{Primaal: } \min \left\{ \sum_{i=1}^{k} (c^i)^T x^i \;\middle|\; \begin{array}{l} \sum_{i=1}^{k} (I - \alpha P^i) x^i = e \\ x^i \geq 0,\ 1 \leq i \leq k \end{array} \right\}$$

$$\text{Duaal: } \max\{e^T y \mid (I - \alpha P^i)^T y + s^i = c^i \text{ en } s^i \geq 0,\ 1 \leq i \leq k\},$$

waarbij $x^i \in \mathbf{R}^n$ correspondeert met de beslissingsvariabelen van alle toestanden voor actie i, en P^i is weer een Markovmatrix.
Als we vergelijken met de standaardvorm van een LP-probleem, dan krijgen we $A = [I - \alpha P^1, \ldots, I - \alpha P^k] \in \mathbf{R}^{n \times nk},\ b = e \in \mathbf{R}^n,$ en $c = (c^1; \ldots; c^k) \in \mathbf{R}^{nk}$.
α is de verdisconteringsfactor zdd. $\alpha = \frac{1}{1+\rho}$ met $\rho > 0$ de rente.
Het probleem is nu om voor elke toestand de beste actie te vinden zdd. de totale kosten minimaal zijn. In dit hoofdstuk wordt een IPM ontwikkeld, waarmee het MBP in maximaal $O(n^{1.5}(\log \frac{1}{1-\alpha} + \log n))$ iteraties en $O(n^4(\log \frac{1}{1-\alpha} + \log n))$ rekenoperaties kan worden opgelost. Het is gebaseerd op het artikel van Ye, september 2002[2]. Tot zover bekend is dit het eerste streng polynomiale algoritme (de rekentijd is onafhankelijk van c en P) voor het oplossen van het MBP als $0 < \alpha < 1$ en α is constant.

2.3 LP-stellingen en inwendigpuntalgoritmes

2.3.1 Optimaliteitsvoorwaarden en het centrale pad

De optimaliteitsvoorwaarden voor een optimale oplossing van $LP(A, b, c)$ zijn (zie ook (1.17) met $\mu = 0$):

$$\begin{array}{rcl} Ax & = & b, \\ A^T y + s & = & c, \\ SXe & = & 0, \\ x \geq 0, & & s \geq 0 \end{array} \tag{2.3}$$

waarbij $X = \text{diag}(x)$ en $S = \text{diag}(s)$. Als $LP(A, b, c)$ een optimale oplossing heeft, dan is er een unieke indexverzameling $B^* \subset \{1, \ldots, n\}$ en $N^* = \{1, \ldots, n\} \setminus B^*$ zdd. iedere x die voldoet aan

$$A_{B^*} x_{B^*} = b,\ x_{B^*} \geq 0,\ x_{N^*} = 0,$$

optimaal is voor het primale probleem; en iedere (y, s) die voldoet aan

$$s_{B^*} = c_{B^*} - A_{B^*}^T y = 0,\ s_{N^*} = c_{N^*} - A_{N^*}^T y \geq 0,$$

is optimaal voor het duale probleem.

De vergelijkingen

$$\begin{array}{rcl} Ax & = & b, \\ A^T y + s & = & c, \\ SXe & = & \mu e, \\ x > 0, & & s > 0 \end{array} \tag{2.4}$$

hebben altijd een unieke oplossing voor alle $\mu > 0$, als zowel het primale als het duale probleem toelaatbare inwendige punten heeft ($x > 0,\ s > 0$). De oplossing

2. Y. Ye, A New Complexity Result on Solving the Markov Decision Problem, 2002 (niet gepubliceerd)

van deze vergelijkingen, $(x(\mu), y(\mu), s(\mu))$, heet het *punt op het centrale pad* voor μ; de verzameling van alle punten, als μ van 0 tot ∞ loopt, is het *centrale pad* van het LP-probleem.
Als $\mu \to 0^+$, dan nadert (2.4) naar (2.3), en $(x(\mu), y(\mu), s(\mu))$ nadert naar een optimale oplossing.
Het centrale pad heeft de volgende meetkundige eigenschap:

Lemma 2.1 *Laat $(x(\mu), y(\mu), s(\mu))$ en $(x(\mu'), y(\mu'), s(\mu'))$ twee punten op het centrale pad zijn zdd. $0 \leq \mu' < \mu$. Dan geldt voor alle i dat*

$$s(\mu')_i \leq n s(\mu)_i \text{ en } x(\mu')_i \leq n x(\mu)_i.$$

I.h.b. geldt bij willekeurige optimale (x^*, y^*, s^*), $\mu > 0$ en i, dat

$$s_i^* \leq n s(\mu)_i \text{ en } x_i^* \leq n x(\mu)_i.$$

Bewijs
Het bewijs staat in [Vavasis]. ♢

2.3.2 Het predictor-corrector-inwendigpuntalgoritme

Bij de predictor-corrector-padvolgende IPM van Mizuno e.a.[3] wordt (2.4) bij benadering opgelost; dit levert een benadering (x, y, s, μ) op zdd.

$$\eta(x, y, s, \mu) := \|SXe/\mu - e\| \leq \eta_0 \tag{2.5}$$

In elke iteratie daalt μ verder naar 0, en (x, y, s, μ) wordt opnieuw berekend.
In [Vavasis] komt het volgende lemma voor:

Lemma 2.2 *Stel dat $\eta(x, y, s, \mu) = \eta < 1$, en laat $(x(\mu), y(\mu), s(\mu))$ het punt op het centrale pad zijn voor parameter μ. Dan geldt voor iedere i:*

$$\left(1 - \frac{\eta}{1-\eta}\right) s_i \leq s_i(\mu)_i \leq \left(1 + \frac{\eta}{1-\eta}\right) s_i$$

en

$$\left(1 - \frac{\eta}{1-\eta}\right) x_i \leq x_i(\mu)_i \leq \left(1 + \frac{\eta}{1-\eta}\right) x_i.$$

Bewijs
Het bewijs staat in [Gonzaga][4]. ♢

Voor $\eta = 1/4$ wordt dit:

$$\begin{aligned} (2/3)s_i \leq s(\mu)_i \leq (4/3)s_i \\ (2/3)x_i \leq x(\mu)_i \leq (4/3)x_i. \end{aligned} \tag{2.6}$$

3. S. Mizuno, M.J. Todd en Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, *Mathematics of Operations Research* **18**, p. 964-981, 1993
4. C.C. Gonzaga, Path-following methods for linear programming, *SIAM-review* **34**, p. 167-224, 1992

Uit (2.5) volgt dat $\mu(1-\eta_0) \leq [SX]_{ii} \leq \mu(1+\eta_0)$, en dus geldt dat

$$\begin{array}{c} \sqrt{\mu(1-\eta_0)} \leq \|(SX)^{1/2}\| \leq \sqrt{\mu(1+\eta_0)} \\ \frac{1}{\sqrt{\mu(1+\eta_0)}} \leq \|(SX)^{-1/2}\| \leq \frac{1}{\sqrt{\mu(1-\eta_0)}}. \end{array} \tag{2.7}$$

De nieuwe μ wordt nu berekend door in een predictor-stap twee gerelateerde kleinste kwadraten problemen (KK-problemen) op te lossen:

$$\min\{\|D^{-1/2}(\delta s + s)\| \mid \delta s = -A^T \delta y, \text{ ofwel } \delta s \in \mathcal{R}(A^T)\},$$

waarbij $\mathcal{R}(A^T)$ de kolomruimte van A^T is, en $D = X^{-1}S$, en

$$\min\{\|D^{1/2}(\delta x + x)\| \mid A\delta x = 0, \text{ ofwel } \delta x \in \mathcal{N}(A)\},$$

waarbij $\mathcal{N}(A)$ de nulruimte van A is. Deze twee KK-problemen hebben oplossingen $(\delta\bar{y}, \delta\bar{s})$ resp. $\delta\bar{x}$.

Lemma 2.3 *Er geldt dat*

$$\begin{array}{rcl} A\delta\bar{x} & = & 0, \\ A^T\delta\bar{y} + \delta\bar{s} & = & 0, \\ D^{1/2}\delta\bar{x} + D^{-1/2}\delta\bar{s} & = & -X^{1/2}S^{1/2}e. \end{array} \tag{2.8}$$

Bewijs
De eerste twee regels van (2.8) volgen direct uit de KK-problemen.
Bij het eerste KK-probleem geldt voor de Lagrangefunctie F dat

$$F(\delta s, \delta y, v) = \frac{1}{2}(\delta s + s)^T D^{-1}(\delta s + s) - v^T(\delta s + A^T\delta y - 0).$$

$\frac{\partial F}{\partial \delta y} = 0 \Rightarrow Av = 0$;
$\frac{\partial F}{\partial v} = 0 \Rightarrow \delta s + A^T\delta y = 0 \Rightarrow \delta s = -A^T\delta y$
$\frac{\partial F}{\partial \delta s} = 0 \Rightarrow D^{-1}(\delta s + s) = v \Rightarrow AD^{-1}(\delta s + s) = Av = 0 \Rightarrow A(\delta s + s) = 0 \Rightarrow -AA^T\delta y + As = 0 \Rightarrow \delta y = (AA^T)^{-1}As$.
Hieruit volgt dat

$$\delta\bar{s} = -A^T(AA^T)^{-1}As.$$

Voor het tweede KK-probleem geldt dat

$$F(\delta x, u) = \frac{1}{2}(\delta x + x)^T D(\delta x + x) - u^T(A\delta x - 0).$$

$\frac{\partial F}{\partial u} = 0 \Rightarrow A\delta x = 0 \Rightarrow AD\delta x = 0$
$\frac{\partial F}{\partial \delta x} = 0 \Rightarrow D(\delta x + x) = A^T u \Rightarrow AD(\delta x + x) = AA^T u$.
Omdat $AD\delta x = 0$, volgt hieruit dat $u = (AA^T)^{-1}ADx$, dus

$$\delta\bar{x} + x = A^T(AA^T)^{-1}Ax.$$

Hieruit volgt dat

$$\begin{array}{rcl} S\delta\bar{x} + X\delta\bar{s} & = & S\{A^T(AA^T)^{-1}AXe - Xe\} - X\{A^T(AA^T)^{-1}ASe\} \\ & = & SX\{A^T(AA^T)^{-1}Ae - A^T(AA^T)^{-1}Ae\} - SXe \\ & = & -SXe. \end{array}$$

$\diamondsuit$

Uit de derde regel van (2.8) volgt dat

$$S\delta\bar{x} + X\delta\bar{s} = -Xs \tag{2.9}$$

Omdat $\delta\bar{x} \in \mathcal{N}(A)$ en $\delta\bar{s} \in \mathcal{R}(A^T)$, is $(\delta\bar{x})^T \cdot \delta\bar{s} = 0$, en hieruit volgt weer dat

$$\|D^{1/2}\delta\bar{x}\|^2 + \|D^{-1/2}\delta\bar{s}\|^2 = n\mu. \tag{2.10}$$

De volgende benadering $(\bar{x}, \bar{y}, \bar{s})$ wordt nu gedefinieerd door

$$(\bar{x}, \bar{y}, \bar{s}) = (x, y, s) + \theta(\delta\bar{x}, \delta\bar{y}, \delta\bar{s}),$$

en dit is voor een geschikte $\theta \in (0, 1]$ een strikt toelaatbaar punt. (Dit noemen we de predictorstap.)

Met de notatie $\Delta\bar{x} = diag(\delta\bar{x})$ en $\Delta\bar{s} = diag(\delta\bar{s})$ geldt verder dat

$$\begin{aligned}
\bar{S}\bar{X} &= (S + \theta\Delta\bar{s})(X + \theta\Delta\bar{x}) \\
&= SX + \theta\{X\Delta\bar{s} + S\Delta\bar{x}\} + \theta^2\Delta\bar{s}\Delta\bar{x} \\
&= SX - \theta XS + \theta^2\Delta\bar{s}\Delta\bar{x} \\
&= (1-\theta)SX + \theta^2\Delta\bar{s}\Delta\bar{x}.
\end{aligned}$$

En dus is

$$\begin{aligned}
\eta(\bar{x}, \bar{y}, \bar{s}, \mu(1-\theta)) &= \frac{1}{\mu(1-\theta)}\|\bar{S}\bar{X}e - \mu(1-\theta)e\| \\
&= \frac{1}{\mu}\|SXe - \mu e + \frac{\theta^2}{1-\theta}\Delta\bar{s}\delta\bar{x}\| \\
&\leq \eta_0 + \frac{\theta^2}{\mu(1-\theta)}\|\Delta\bar{s}\delta\bar{x}\|.
\end{aligned}$$

Het is dus mogelijk om θ zo te kiezen dat

$$\eta(\bar{x}, \bar{y}, \bar{s}, \mu(1-\theta)) \leq 2\eta_0.$$

Vervolgens wordt een correctorstap genomen: m.b.v.

$$\begin{aligned}
\bar{S}\delta\bar{x} + \bar{X}\delta\bar{s} &= \mu e - \bar{X}\bar{s} \\
A\delta\bar{x} &= 0, \\
-A^T\delta\bar{y} - \delta\bar{s} &= 0
\end{aligned}$$

wordt vanuit $(\bar{x}, \bar{y}, \bar{s})$ een nieuwe (x, y, s) berekend zdd. $\eta(x, y, s, \mu(1-\theta)) \leq \eta_0$. (Zie [Ye, 1997][5], p. 131 en (4.17)). Het predictor-corrector-inwendigpunt algoritme wordt nu:

1. Start met (x, y, s, μ) zdd. $\eta(x, y, s, \mu) \leq \eta_0 = 1/4$. Kies $\epsilon > 0$.
2. Predictorstap: bereken $(\delta\bar{x}, \delta\bar{y}, \delta\bar{s})$ m.b.v. (2.8);
   $(\bar{x}, \bar{y}, \bar{s}) := (x, y, s) + \bar{\theta}(\delta\bar{x}, \delta\bar{y}, \delta\bar{s})$, waarbij $\bar{\theta} \in (0, 1)$ de grootste θ is zdd. $\eta(\bar{x}, \bar{y}, \bar{s}, \mu(1-\bar{\theta})) \leq 2\eta_0$.

5. Y. Ye, *Interior Point Algorithms: Theory and Analysis*, JohnWiley & Sons, 1997

3. Correctorstap: bereken $(\delta\bar{x}', \delta\bar{y}', \delta\bar{s}')$ m.b.v.

$$\begin{aligned} \bar{S}\delta\bar{x}' + \bar{X}\delta\bar{s}' &= \mu e - \bar{X}\bar{s} \\ A\delta\bar{x}' &= 0, \\ -A^T\delta\bar{y}' - \delta\bar{s}' &= 0; \end{aligned}$$

   $(x, y, s) := (\bar{x}, \bar{y}, \bar{s}) + (\delta\bar{x}', \delta\bar{y}', \delta\bar{s}')$.
4. Als $x^T s \leq \epsilon$: STOP;
   Anders: ga naar stap 2.

Voor de stapgrootte θ en voor het aantal iteraties gelden de volgende lemma's:

Lemma 2.4 *De stapgrootte kan in iedere iteratie groter zijn dan* $\frac{1}{4\sqrt{n}}$.

Bewijs
In [Mizuno] staat (Lemma 4, p. 971): *Als* $\eta_0 = 1/4$, *dan voldoet de stapgrootte in de predictorstap aan*

$$\theta \geq \theta_1 := \min\left\{\frac{1}{2}, \left(\frac{\mu}{8\|Pq\|}\right)^{1/2}\right\},$$

waarbij $p := X^{-1/2}S^{1/2}\delta\bar{x}$, $q := X^{1/2}S^{-1/2}\delta\bar{s}$ en $P := \mathrm{diag}(p)$. Uit Lemma 4.14(i) en Lemma 4.15(i) uit [Ye,1997] volgt dat $\|Pq\| \leq \frac{\sqrt{2}}{4}n\mu$, en hieruit volgt dat

$$\theta \geq \min\left\{\frac{1}{2}, \left(\frac{\mu}{2\sqrt{2}n\mu}\right)^{1/2}\right\} = \min\left\{\frac{1}{2}, \frac{1}{8^{1/4}\sqrt{n}}\right\} = \frac{1}{2^{3/4}\sqrt{n}} > \frac{1}{4\sqrt{n}}$$

als $n \geq 2$. ♢

Lemma 2.5 *Het predictor-corrector-inwendigpuntalgoritme reduceert* μ *tot* μ' $(<$ $\mu)$ *in maximaal* $O(\sqrt{n}\log(\mu/\mu'))$ *iteraties.*

Bewijs
In [Ye, 1997] staat (Stelling 4.18): *als* $\eta_0 = 1/4$, *dan stopt algoritme 4.5 na* $k \leq O(\sqrt{n}\log((x^0)^T s^0/\epsilon))$ *iteraties.*
Dan is dus $\mu^k := (x^k)^T s^k \leq \epsilon$.
Dus het algoritme reduceert $\mu := (x^0)^T s^0$ tot $\mu' := \mu^k$ in k iteraties. Omdat $\epsilon \geq (x^k)^T s^k = \mu$, is $O(\sqrt{n}\log((x^0)^T s^0/\epsilon)) \leq O(\sqrt{n}\log(\mu/\mu'))$. ♢

2.3.3 Eigenschappen van het MBP

We nemen vanaf nu $k = 2$. Dan wordt het MBP:

$$\min\left\{(c^1)^T x^1 + (c^2)^T x^2 \;\middle|\; \begin{array}{l} (I - \alpha P^1)x^1 + (I - \alpha P^2)x^2 = e \\ x^1,\ x^2 \geq 0 \end{array}\right\} \tag{2.11}$$

en het duale probleem wordt:

$$\max\left\{e^T y \;\middle|\; \begin{array}{l} (I - \alpha P^1)^T y + s^1 = c^1 \\ (I - \alpha P^2)^T y + s^2 = c^2 \\ s^1,\ s^2 \geq 0 \end{array}\right\} \tag{2.12}$$

Lemma 2.6 *Het MBP heeft de volgende eigenschappen:*
(i) Het primale en duale MBP hebben inwendige toelaatbare punten als $0 \leq \alpha < 1$.
(ii) De verzameling toelaatbare punten van het primale MBP is begrensd:

$$e^T x = \frac{n}{1-\alpha},$$

met $x = (x^1; x^2)$.
(iii) Laat $\hat{x}$ *een toelaatbare basisoplossing van het MBP zijn. Dan geldt voor iedere basisvariabele* $\hat{x}_i$ *dat* $\hat{x}_i \geq 1$.
(iv) Laat B^* *en* N^* *de optimale opsplitsing van het MBP zijn. Dan bevat* B^* *minstens één toelaatbare basis:* $|B^*| \geq n$ *en* $|N^*| \leq n$;
voor alle $j \in B^*$ *is er een optimale oplossing* x^* *zdd.* $x^*_j \geq 1$.
(v) Laat A_B *een willekeurige toelaatbare basis zijn en* A_N *een willekeurige deelmatrix van de restkolommen van de matrix A. Dan is*

$$\|(A_B)^{-1} A_N\| \leq \frac{2n\sqrt{n}}{1-\alpha}.$$

Bewijs
Het bewijs staat in [Ye, 2002]. Er wordt gebruikt dat $e^T P = e^T$, $(I - \alpha P)^{-1} = I + \alpha P + \alpha^2 P^2 + \ldots$, en het feit dat P^k een Markovmatrix blijft voor $k = 1, 2, \ldots$.
$\diamondsuit$

We mogen aannemen dat $c \geq 0$, omdat de optimale oplossing van het MBP niet verandert als we $c + \gamma e$ i.p.v. c gebruiken.

2.4 Combinatorische inwendigpuntmethode voor het oplossen van het MBP

De methode die wordt besproken, heeft maximaal n hoofdstappen, waarin in iedere stap minstens één variabele in N^* wordt geëlimineerd. De methode gaat dan verder met minstens één variabele minder. Elke hoofdstap gebruikt maximaal $O(n^{0.5}(\log \frac{1}{1-\alpha} + \log n))$ predictor-corrector-iteraties.

2.4.1 Complexiteit om een punt dicht bij het centrale pad te berekenen

Voor de methode is eerst een punt (x, y, s, μ) dicht bij het centrale pad nodig dat voldoet aan $\|SXe/\mu - e\| \leq \eta_0$. Neem

$$(x^i)^0 = (I - \alpha P^i)^{-1} e, \; i = 1, 2$$

en

$$x^0 = \begin{pmatrix} \frac{1}{2}(x^1)^0 \\ \frac{1}{2}(x^2)^0 \end{pmatrix}.$$

Dan is x^0 een inwendig toelaatbaar punt voor het MBP, en

$$x^0 \geq \frac{1}{2} e \in \mathbf{R}^{2n}.$$

Neem verder

$$y^0 = -\gamma e \text{ en } s^0 = \begin{pmatrix} (s^1)^0 \\ (s^2)^0 \end{pmatrix} = \begin{pmatrix} c^1 + \gamma(1-\alpha)e \\ c^2 + \gamma(1-\alpha)e \end{pmatrix}$$

waarbij γ groot genoeg wordt gekozen zdd.

$$s^0 > 0 \text{ en } \gamma \geq \frac{c^T x^0}{n}.$$

Neem $\mu^0 = (x^0)^T s^0/(2n)$ en beschouw de potentiaalfunctie

$$\phi(x,s) := 2n \log(s^T x) - \sum_{j=1}^{2n} \log(s_j x_j).$$

Laat c_j de j-de component zijn van $c = (c^1; c^2) \geq 0$. Dan geldt na enig rekenwerk, zie [Ye, 2002], dat

$$\phi(x^0, s^0) \leq 2n \log(2n) + 2n \log\left(\frac{2}{1-\alpha}\right).$$

Met dit algortime kunnen we dus (volgens [Ye, 2002]) een (x^0, y^0, s^0) vinden zdd. $\eta(x^0, y^0, s^0, \mu^0) \leq \eta_0$. Hiervoor zijn maximaal $O(n(\log \frac{2}{1-\alpha} + \log n))$ iteraties nodig, en per iteratie $O(n^3)$ rekenkundige bewerkingen.

2.4.2 Het elimineren van variabelen in N^*

Lemma 2.7 *Voor willekeurige $\mu \in (0, \mu^0]$ voldoen de oplossingen $x(\mu)$ en $s(\mu)$ van (2.11) en (2.12) aan*

$$x(\mu)_j \leq \frac{n}{1-\alpha} \text{ en } s(\mu)_j \geq \frac{1-\alpha}{n}\mu \ \forall j = 1, \ldots, 2n,$$

en

$$x(\mu)_j \geq \frac{1}{2n} \text{ en } s(\mu)_j \leq 2n\mu \ \forall j \in B^*.$$

Bewijs
Volgens Lemma 2.6(ii) geldt dat $e^T x = \frac{n}{1-\alpha}$. Hieruit volgt dat $x(\mu)_j \leq \frac{n}{1-\alpha}$ $\forall j = 1, \ldots, 2n$.
Omdat er $2n$ variabelen zijn, volgt uit Lemma 2.1 en 2.6(iv) dat $2n x(\mu)_j \geq x^*_j \geq 1$ $\forall j \in B^*$.
De grenzen voor $s(\mu)_j$ volgen uit (2.4). ♢

Definieer een *klooffactor*

$$g := \frac{10n^2(1+\eta_0)}{(1-\alpha)\sqrt{1-\eta_0}} \ (> 1), \tag{2.13}$$

en definieer voor een willekeurig punt (x, y, s) in de buurt van het centrale pad zdd. $\eta(x, y, s, \mu) \leq \eta_0$:

$$\begin{aligned} J_1(\mu) &= \{j \mid s_j \leq 3n\mu\}, \\ J_3(\mu) &= \{j \mid s_j \geq 3n\mu \cdot g\}, \text{ en} \\ J_2(\mu) &= \{\text{de rest van de indices}\}. \end{aligned}$$

Dan geldt voor alle $j_1 \in J_1(\mu)$ en $j_3 \in J_3(\mu)$ dat

$$\frac{s_{j_1}}{s_{j_3}} \leq \frac{1}{g} < 1. \tag{2.14}$$

Uit lemma 2.7 en (2.6) volgt nu voor willekeurige $j \in B^*$ dat

$$s_j = \frac{s_j}{s(\mu)_j} s(\mu)_j \leq \frac{s_j}{s(\mu)_j} 2n\mu \leq \frac{3}{2} 2n\mu = 3n\mu,$$

en hieruit volgt:

Lemma 2.8 *Met de definities voor $J_1(\mu)$ en $J_3(\mu)$ geldt:*
(i) $B^* \subset J_1(\mu)$ $\forall \mu$ zdd. $0 < \mu \leq \mu_0$, dus $J_1(\mu)$ bevat altijd een optimale basis.
(ii) Omdat $g > 1$, is $J_3(\mu) \subset N^*$.

Omdat $J_3(\mu) \subset N^*$, is iedere primale variabele in $J_3(\mu)$ gelijk aan 0 in iedere primale optimale oplossing. Dus kunnen we, als $J_3(\mu) \neq \emptyset$, elke primale variabele in $J_3(\mu)$ verder buiten beschouwing laten (elimineren). Om toelaatbaarheid te garanderen, lossen we het volgende KK-probleem op:

$$\min_{\delta x_1} \left\{ \ \|D_1^{1/2} \delta x_1\| \ \middle| \ A_1 \delta x_1 = A_3 x_3 \ \right\} \tag{2.15}$$

Index i geeft hier de deelvector of deelmatrix van indexverzameling $J_i(\mu)$ aan, en $D_1 = X_1^{-1} S_1$. Dit probleem is altijd toelaatbaar, want A_1 bevat B^* en B^* bevat minstens één optimale basis.
Er geldt dan dat

$$A_1(x_1 + \delta x_1) + A_2 x_2 = A_1 x_1 + A_2 x_2 + A_3 x_3 = b.$$

De vraag is nu of $x_1 \delta x_1 > 0$:

Lemma 2.9 *Er geldt dat $A_1(x_1 + \delta x_1) + A_2 x_2 = b$ en $(x_1 + \delta x_1; x_2) > 0$, en, als $\eta_0 \geq \frac{1+\sqrt{101}}{50}$, ook dat*

$$\eta((x_1 + \delta x_1; x_2), y, (s_1; s_2), \mu) \leq 2\eta_0.$$

M.a.w., het is een punt dicht bij het centrale pad voor (2.11) en (2.12) voor dezelfde μ, na het verwijderen van alle primale variabelen en duale voorwaarden in $J_3(\mu)$.

Bewijs
Laat A_B een optimale basis zijn met $A_B \subset A_1$, en laat B de indexverzameling van de basis zijn. Dan geldt dat (zie hieronder)

$$A_1 D_1^{-1} A_1^T \succeq A_B D_B^{-1} A_B^T \succ 0,$$

waarbij betekent $U \succeq V$ dat $U - V$ positief semi-definiet is, en $V \succ 0$ betekent dat V positief definiet is:
Schrijf

$$A_1 = \left(\ A_B \ \middle| \ R \ \right) \text{ en } D_1 = \left(\begin{array}{c|c} D_B & 0 \\ \hline 0 & D_R \end{array} \right).$$

Dan is

$$A_1 D_1^{-1} A_1^T = \left(\begin{array}{c|c} A_B D_B^{-1} & R D_R^{-1} \end{array} \right) \left(\frac{A_B^T}{R^T} \right) = A_B D_B^{-1} A_B^T + R D_R^{-1} R^T.$$

Dus $A_1 D_1^{-1} A_1^T - A_B D_B^{-1} A_B^T = R D_R^{-1} R^T$ en deze matrix is positief semi-definiet, want $D_R \geq 0$. Het KK-probleem (2.15) is equivalent met

$$\min_{\delta x_1} \left\{ \begin{array}{c|c} \frac{1}{2} \|D_1^{1/2} \delta x_1\|^2 & A_1 \delta x_1 = A_3 x_3 \end{array} \right\},$$

en hierbij hoort de Lagrangefunctie

$$F(\delta x_1, u) = \frac{1}{2} \delta x_1^T D_1 \delta x_1 - u^T (A_1 \delta x_1 - A_3 x_3).$$

$\frac{\partial F}{\partial u} = 0 \Rightarrow A_1 \delta x_1 = A_3 x_3$.
$\frac{\partial F}{\partial \delta x_1} = 0 \Rightarrow D_1 \delta x_1 = A_1^T u \Rightarrow A_1 D_1 \delta x_1 = A_1 A_1^T u \Rightarrow A_1 \delta x_1 = A_1 D_1^{-1} A_1^T u \Rightarrow A_3 x_3 = A_1 D_1^{-1} A_1^T u \Rightarrow u = (A_1 D_1^{-1} A_1^T)^{-1} A_3 x_3$, en dus is

$$\delta x_1 = D_1^{-1} A_1^T u = D_1^{-1} A_1^T (A_1 D_1^{-1} A_1^T)^{-1} A_3 x_3.$$

Hieruit volgt, zie [Ye, 2002], dat

$$\|D_1^{1/2} \delta x_1\| \leq \frac{\sqrt{\mu}}{5},$$

en

$$\|X_1^{-1} \delta x_1\| < 1.$$

Dus

$$x_1 + \delta x_1 = X_1 (e + X_1^{-1} \delta x_1) > 0.$$

Verder is, zie [Ye, 2002],

$$\left\| \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 \end{pmatrix} - \mu e \right\| \leq \eta_0 \mu + \sqrt{1 + \eta_0} \mu / 5,$$

en als $\eta_0 \geq \frac{1+\sqrt{101}}{50} \approx 0.221$, dan is $\eta_0 \mu + \sqrt{1 + \eta_0} \mu / 5 \leq 2 \eta_0 \mu$. ♢

2.5 Complexiteit om ervoor te zorgen dat $J_3(\mu) \neq \emptyset$

Als $J_3(\mu) = \emptyset$ bij beginwaarde μ^0, dan passen we de predictor-corrector-methode toe: we berekenen de predictorstap m.b.v. (2.8) in het punt (x, y, s) met $\eta(x, y, s, \mu^0) \leq \eta_0$.
We bekijken eerst het geval dat $N^* = \emptyset$.

Lemma 2.10 *Als $N^* = \emptyset$, dan is iedere toelaatbare oplossing van (2.11) een optimale oplossing.*

Bewijs
Als $N^* = \emptyset$, dan is $s = c - A^T y = 0$, dus $x^T s = 0$. Dus iedere toelaatbare oplossing van (2.11) is optimaal. ♢

We mogen daarom aannemen dat

$$\epsilon^0 := \frac{1}{\sqrt{\mu^0}}\|D^{-1/2}(\delta\bar{s} + s)\| = \frac{1}{\sqrt{\mu^0}}\|D^{1/2}\delta\bar{x}\| > 0. \tag{2.16}$$

Definieer verder

$$\bar{\theta} = \max\left\{0, 1 - \frac{\sqrt{n}\epsilon^0}{\eta_0}\right\}. \tag{2.17}$$

Neem nu een stap waarvan de richting wordt bepaald door (2.8); dit levert de volgende $(\bar{x}, \bar{y}, \bar{s})$ op:

$$\begin{aligned} \bar{x} &= x + \bar{\theta}\delta\bar{x}, \\ \bar{y} &= y + \bar{\theta}\delta\bar{y}, \\ \bar{s} &= s + \bar{\theta}\delta\bar{s}. \end{aligned}$$

Lemma 2.11 *Als $\bar{\theta} < 1$ in (2.17), dan is $(\bar{x}, \bar{y}, \bar{s})$ strikt toelaatbaar. Bovendien is*

$$\eta(\bar{x}, \bar{y}, \bar{s}, \mu^0(1 - \bar{\theta})) \leq 2\eta_0.$$

Bewijs
Het lemma is waar voor $\bar{\theta} = 0$. Neem daarom aan $\bar{\theta} > 0$:

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} < 1 \text{ ofwel } \bar{\theta} = 1 - \frac{\sqrt{n}\epsilon^0}{\eta_0} > 0.$$

Laat $\theta \in (0, \bar{\theta}]$ en definieer

$$\begin{aligned} x(\theta) &= x + \theta\delta\bar{x} \\ y(\theta) &= y + \theta\delta\bar{y} \\ s(\theta) &= s + \theta\delta\bar{s} \end{aligned}$$

Dan is m.b.v. (2.9) en (2.10) te bewijzen dat

$$\eta(x(\theta), y(\theta), s(\theta), \mu^0(1 - \theta)) \leq 2\eta_0,$$

zie [Ye, 2002].
Nu rest het bewijs dat $(x(\theta), y(\theta), s(\theta))$ toelaatbaar is. De afstandsmaat $\eta(x(\theta), y(\theta), s(\theta), \mu^0(1 - \theta)) = \|S(\theta)X(\theta)e/(\mu^0(1 - \theta)) - e\|$ is een continue functie van θ voor $\theta \in (0, \bar{\theta}]$. Hierboven is bewezen dat $\eta(x(\theta), y(\theta), s(\theta), \mu^0(1 - \theta)) \leq 2\eta_0 < 1\ \forall\theta \in (0, \bar{\theta}]$. Dit kan alleen als $x(\theta)_i, s(\theta)_i \neq 0\ \forall i = 1, \ldots, 2n$. Ook geldt dat $x(0) = x > 0$ en $s(0) = s > 0$, en dus is $s(\theta) > 0$ en $x(\theta) > 0$ voor alle $\theta \in (0, \bar{\theta}]$. ♢

Lemma 2.12 *Als $\epsilon^0 > 0$, dan is er een variabele met index $\bar{j}$ zdd. $\bar{j} \in N^*$, en*

$$s(\mu)_{\bar{j}} \geq \frac{\sqrt{1 - \eta_0}(1 - \alpha)\mu^0}{2\sqrt{2}n^{2.5}}\epsilon^0$$

voor alle $\mu \in (0, \mu^0]$.

Bewijs
Er geldt (zie [Ye, 2002]) dat

$$\|s^*\|_\infty \geq \frac{\sqrt{1-\eta_0}(1-\alpha)\mu^0}{\sqrt{2}n^{1.5}} \cdot \epsilon^0.$$

Uit Lemma 2.1 (toegepast op $2n$ variabelen) volgt dat er een variabele met index $\bar{j}$ is, zdd. $\bar{j} \in N^*$, en

$$s(\mu)_{\bar{j}} \geq \frac{s^*_{\bar{j}}}{2n} = \frac{\|s^*\|_\infty}{2n} \geq \frac{\sqrt{1-\eta_0}(1-\alpha)\mu^0}{2\sqrt{2}n^{2.5}} \ \forall \mu \in (0, \mu^0].$$

$\diamondsuit$

Er zijn nu twee mogelijkheden:

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} \geq 1 \tag{2.18}$$

en

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} < 1. \tag{2.19}$$

Als (2.18) geldt, dan is $\bar{\theta} = 0$ en

$$\epsilon^0 \geq \frac{\eta_0}{\sqrt{n}} \text{ en } s(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\alpha)\mu^0}{2\sqrt{2}n^3},$$

waarbij index $\bar{j} \in N^*$ de index uit lemma 2.12 is. Pas in dit geval het predictor-corrector-padvolgende algoritme toe totdat

$$\frac{\mu}{\mu^0} \leq \frac{\eta_0\sqrt{1-\eta_0}(1-\alpha)}{8\sqrt{2}n^4 \cdot g}.$$

Dan is

$$s(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\alpha)\mu^0}{2\sqrt{2}n^3} \geq 4n\mu \cdot g,$$

en in een willekeurige (x, y, s) zdd. $\eta(x, y, s, \mu) \leq \eta_0$, is

$$s_{\bar{j}} \geq \frac{4}{5}s(\mu)_{\bar{j}} \geq \frac{16}{5}n\mu \cdot g > 3n\mu \cdot g.$$

Dus nu is $\bar{j} \in J_3(\mu)$ en hij kan geëlimineerd worden.
Als (2.19) geldt, dan is

$$1 - \bar{\theta} = \frac{\sqrt{n}\epsilon^0}{\eta_0} \text{ en } s(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\alpha)(1-\bar{\theta})\mu^0}{2\sqrt{2}n^3},$$

waarbij index $\bar{j} \in N^*$ weer de index uit lemma 2.12 is. Na één predictorstap wordt μ^0 gereduceerd tot $(1-\bar{\theta})\mu^0$. Pas daarna weer het predictor-corrector-padvolgende algoritme toe totdat

$$\frac{\mu}{(1-\bar{\theta})\mu^0} \leq \frac{\eta_0\sqrt{1-\eta_0}(1-\alpha)}{8\sqrt{2}n^4 \cdot g}.$$

Dan is

$$s(\mu)_{\bar{j}} \geq 4n\mu \cdot g,$$

en in een willekeurige (x, y, s) zdd. $\eta(x, y, s, \mu) \leq \eta_0$, is

$$s_{\bar{j}} \geq \frac{4}{5} s(\mu)_{\bar{j}} \geq \frac{16}{5} n\mu \cdot g > 3n\mu \cdot g.$$

Dus nu is $\bar{j} \in J_3(\mu)$ en hij kan geëlimineerd worden.

In beide gevallen zijn maximaal $O(n^{0.5}(\log \frac{1}{1-\alpha} + \log n))$ iteraties van het predictor-corrector-algoritme nodig om μ voldoende te reduceren:
na één iteratie van het predictor-corrector-inwendigpuntalgoritme wordt μ een factor $1 - \theta$ kleiner, en het algoritme reduceert μ tot μ' in maximaal $O(\sqrt{n} \log \frac{\mu}{\mu'})$ iteraties.
In geval (2.18) geldt: zodra $\frac{\mu'}{\mu} \leq \frac{\eta_0(1-\eta_0)(1-\alpha)}{80\sqrt{2}n^6(1+\eta_0)}$, kan $\bar{j}$ geëlimineerd worden. Dus, als $\eta_0 \in [1/5, 1/4]$, dan is

$$\frac{\mu'}{\mu(1-\theta)} > \frac{\eta_0(1-\eta_0)(1-\alpha)}{80\sqrt{2}n^6(1+\eta_0)} \Rightarrow$$

$$\frac{\mu}{\mu'} < \frac{80\sqrt{2}n^6(1+\eta_0)}{\eta_0(1-\eta_0)(1-\alpha)(1-\theta)} \leq \frac{80\sqrt{2}n^6 \cdot 15/2}{(1-\alpha)(1-\theta)} = \frac{600\sqrt{2}n^6}{(1-\alpha)(1-\theta)}$$

en dit is $\leq \frac{n^6 \cdot \text{constante}}{(1-\alpha)}$ omdat $\theta \in (0, 1)$.
In geval (2.19) geldt: zodra $\frac{\mu'}{\mu(1-\theta)} \leq \frac{\eta_0(1-\eta_0)(1-\alpha)}{80\sqrt{2}n^6(1+\eta_0)}$, kan $\bar{j}$ geëlimineerd worden. Dus:

$$\frac{\mu'}{\mu(1-\bar{\theta})(1-\theta)} > \frac{\eta_0(1-\eta_0)(1-\alpha)}{80\sqrt{2}n^6(1+\eta_0)} \Rightarrow$$

$$\frac{\mu}{\mu'} < \frac{80\sqrt{2}n^6(1+\eta_0)}{\eta_0(1-\eta_0)(1-\alpha)(1-\theta)(1-\bar{\theta})} \leq \frac{600\sqrt{2}n^6}{(1-\alpha)(1-\bar{\theta})^2}$$

en dit is $\leq \frac{n^6 \cdot \text{constante}}{(1-\alpha)}$ omdat $\bar{\theta} \in (0, 1)$.
Er zijn dus maximaal $O(\sqrt{n}(\log \frac{1}{1-\alpha} + 6 \log n + \log \frac{600\sqrt{2}}{(1-\theta)^2})) = O(\sqrt{n}(\log \frac{1}{1-\alpha} + \log n))$ iteraties nodig.

Deze hoofdstap is maximaal $|N^*| \leq n$ keer nodig; uiteindelijk is $N^* = \emptyset$ en stopt het algoritme volgens lemma 2.10.
Samengevat:

Stelling 2.1 *Het combinatorische inwendigpuntalgoritme geeft een optimale oplossing van (2.11) in maximaal n eliminatiestappen, en iedere stap gebruikt $O(n^{0.5} \cdot (\log \frac{1}{1-\alpha} + \log n))$ iteraties van het predictor-corrector-inwendigpuntalgoritme.*

M.b.v. het 'Karmakar rank one updating scheme' zijn er per iteratie gemiddeld $O(n^{2.5})$ rekenkundige operaties nodig. Hieruit volgt:

Stelling 2.2 *Het combinatorische inwendigpuntalgoritme geeft een optimale oplossing van (2.11) in maximaal $O(n^4(\log \frac{1}{1-\alpha} + \log n))$ rekenkundige operaties.*

Volgens [Ye, 2002] is bij een MBP met k acties per toestand op vergelijkbare wijze te bewijzen:

Stelling 2.3 *Het combinatorische inwendigpuntalgoritme geeft een optimale oplossing van het MBP in maximaal* $(k-1)n$ *eliminatiestappen, en iedere stap gebruikt* $O((nk)^{0.5}(\log \frac{1}{1-\alpha}+\log n+\log k))$ *iteraties van het predictor-corrector-inwendigpuntalgoritme, waar* n *het aantal toestanden is en* k *het aantal acties voor iedere toestand. Het MBP is zo op te lossen met in totaal maximaal* $O((nk)^4(\log \frac{1}{1-\alpha}+\log n+\log k))$ *rekenkundige operaties.*

Hoofdstuk 3

Gemiddelde-opbrengst-criterium

In dit hoofdstuk wordt voor de waardevector $v^{\alpha}(\pi^{\infty})$ een Laurentreeks ontwikkeld. Daarna worden extra gevoelige optimalitieitscrtieria ingevoerd. Voor het criterium van *n-verdisconteerde optimaliteit* wordt tenslotte een algoritme gegeven. Dit hoofdstuk is gebaseerd op [Kallenberg 2][1].

3.1 Stationaire, fundamentele en afwijkingsmatrix

Een $N \times N$-matrix P is een *stochastische matrix* als $p_{ij} \geq 0$ en $\sum_j p_{ij} = 1 \ \forall i, j = 1, \ldots, N$. Het is de overgangsmatrix van een stationaire Markovketen met toestandsruimte $\{1, \ldots, N\}$.
$p_{ij}^{(n)} := [P^n]_{ij}$ geeft de kans aan dat het proces in n stappen van toestand i naar toestand j gaat.
Toestand j is *bereikbaar* vanaf i als $p_{ij}^{(n)} > 0$ voor zekere $n \in \mathbf{N}_0$. Notatie: $i \to j$.
Toestand i is *absorberend* als $p_{ii} = 1$, m.a.w. vanuit toestand i is alleen toestand i bereikbaar.
i en j *communiceren* als $i \to j$ en $j \to i$. Notatie: $i \leftrightarrow j$.
i is *recurrent* als $i \to j$ impliceert dat $j \to i$.
R is de verzameling van alle recurrente toestanden. Uit de definitie van recurrentie volgt dat R *gesloten* is, d.w.z. $p_{ij} = 0 \ \forall i \in R, j \notin R$.
$\leftrightarrow$ is een equivalentierelatie, en dus kan R opgesplitst worden in deelverzamelingen: $R = R_1 \cup R_2 \cup \cdots \cup R_m$, waarbij in R_k elk tweetal toestanden communiceert, en er geen echte deelverzameling van R_k bestaat met deze eigenschap.
De deelverzamelingen R_k heten *ergodische verzamelingen*.
Als $m = 1$ en $R = S$, waarbij S de toestandsruimte is, dan wordt de matrix *irreducibel* genoemd.
De niet-recurrente toestanden worden *transiënt* genoemd.
Als de ergodische verzamelingen $R_1, \ldots, R_m$ en de transiënte toestanden T bekend zijn, dan kan de stochastische matrix, na hernummeren van de toestanden, als volgt geschreven worden:

$$P = \begin{pmatrix} P_1 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & P_m & 0 \\ Q_1 & Q_2 & \cdots & Q_m & Q \end{pmatrix} \begin{matrix} R_1 \\ R_2 \\ \vdots \\ R_m \\ T \end{matrix} \tag{3.1}$$

1. L.C.M. Kallenberg, *Markov Decision Theory*, Univ. Leiden

Vorm (3.1) is de *standaardvorm* van een stochastische matrix.
Voor de recurrente toestanden $i \in R$ wordt de *periode* d_i gedefinieerd door

$$d_i := \text{ggd}\{n \in \mathbf{N} \mid p_{ii}^{(n)} > 0\}. \tag{3.2}$$

Er kan bewezen worden dat de toestanden van een ergodische verzameling dezelfde periode hebben (zie bijv. [Kallenberg, 2001]).
Deze periode wordt de *periode van de ergodische verzameling* genoemd.
Een ergodische verzameling met periode 1 wordt *aperiodiek* genoemd.

Diverse eigenschappen van stochastische matrices
Laat P een stochastische matrix in standaardvorm (3.1) zijn, en neem aan dat $T \neq \emptyset$. Volgens [Kallenberg, 2001] geldt dan dat $Q^n \to 0$ als $n \to \infty$. Omdat

$$(I - Q)(I + Q + Q^2 + \ldots + Q^{n-1}) = I - Q^n,\ n \in \mathbf{N}, \tag{3.3}$$

is het rechterlid van (3.3) regulier. Dus ook het linkerlid is regulier, en we krijgen het volgende resultaat (de *Neumann-reeksontwikkeling*):

Lemma 3.1 *$I - Q$ is regulier en $(I - Q)^{-1} = \sum_{n=0}^{\infty} Q^n$.*

Opmerking: voor $i, j \in T$ kan $[\sum_{n=0}^{\infty} Q^n]_{ij}$ geïnterpreteerd worden als het verwachte aantal keren dat toestand j bezocht wordt als in i wordt gestart. Omdat $i, j \in T$, is dit aantal eindig.

Het limietgedrag van P^n als $n \to \infty$.
In het algemeen bestaat $\lim_{n\to\infty} P^n$ niet: neem bijv. $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, dan is $P^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ als n is even, en $P^n = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ als n is oneven. Dus $\lim_{n\to\infty} P^n$ bestaat niet.
Daarom bekijken we twee andere soorten convergentie:

1. De rij $\{B_n\}_{n=0}^{\infty}$ heet *Cesaro-convergent met Cesaro-limiet B* als

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} B_k$$

bestaat en gelijk is aan B. Notatie: $\lim_{n\to\infty} B_n \overset{(C)}{=} B$, of $B_n \overset{(C)}{\to} B$.

2. De rij $\{B_n\}_{n=0}^{\infty}$ heet *Abel-convergent met Abel-limiet B* als

$$\lim_{\alpha\uparrow 1} (1 - \alpha) \sum_{n=0}^{\infty} \alpha^n B_n$$

bestaat en gelijk is aan B. Notatie: $\lim_{n\to\infty} B_n \overset{(A)}{=} B$, of $B_n \overset{(A)}{\to} B$.

De volgende stelling geeft het verband tussen deze twee soorten convergentie aan:

Stelling 3.1 *Als de rij $\{B_n\}_{n=0}^{\infty}$ Cesaro-convergent is met Cesaro-limiet B, dan is $\{B_n\}_{n=0}^{\infty}$ ook Abel-convergent, en $B_n \overset{(A)}{\to} B$.*

Bewijs
Zie [Powell en Shah, 1972][2] of [Widder][3]. ♢

2. R.E. Powell en S.M Shah, *Summability theory and applications*, Van Nostrand Reinhold, Londen, 1972
3. D. Widder, *Laplace transform*, Princeton University Press, Princeton, New Jersey, 1946

Verder geldt het volgende:

Stelling 3.2 *(i) Gewone convergentie impliceert Cesaro-convergentie;*
(ii) Gewone convergentie impliceert Abel-convergentie;
(iii) Het omgekeerde van (i) en (ii) geldt niet;
(iv) Abel-convergentie impliceert geen Cesaro-convergentie.

Bewijs
(i) De rij $\{B_n\}_{n=0}^{\infty}$ is 'gewoon' convergent als

$$\forall \epsilon > 0 \; \exists M \in \mathbf{N} : \; \forall n \geq M : \; |B_n - B| \leq \epsilon.$$

Er geldt dus dat

$$\left| \left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} B_k \right) - B \right| = \left| \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (B_k - B) \right| =$$

$$\left| \lim_{n\to\infty} \left(\frac{1}{n} \sum_{k=0}^{M-1} (B_k - B) + \sum_{k=M}^{n-1} (B_k - B) \right) \right| \leq$$

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{k=0}^{M-1} (B_k - B) \right| + \lim_{n\to\infty} \left| \frac{1}{n} \sum_{k=M}^{n-1} (B_k - B) \right| =$$

$$0 + \lim_{n\to\infty} \left| \frac{1}{n} \sum_{k=M}^{n-1} (B_k - B) \right| \leq \lim_{n\to\infty} \frac{1}{n} \cdot n\epsilon = \epsilon.$$

Dus $B_n \overset{(C)}{\to} B$.
(ii)

$$\left| \left(\lim_{\alpha\uparrow 1} (1-\alpha) \sum_{n=0}^{\infty} \alpha^n B_n \right) - B \right| = \left| \lim_{\alpha\uparrow 1} (1-\alpha) \sum_{n=0}^{\infty} \alpha^n (B_n - B) \right| =$$

$$\left| \lim_{\alpha\uparrow 1} (1-\alpha) \left(\sum_{n=0}^{M-1} \alpha^n (B_n - B) + \sum_{n=M}^{\infty} \alpha^n (B_n - B) \right) \right| =$$

$$\lim_{\alpha\uparrow 1} (1-\alpha) \left| \sum_{n=0}^{M-1} \alpha^n (B_n - B) \right| + \lim_{\alpha\uparrow 1} (1-\alpha) \left| \sum_{n=M}^{\infty} \alpha^n (B_n - B) \right| =$$

$$0 + \lim_{\alpha\uparrow 1} (1-\alpha) \left| \sum_{n=M}^{\infty} \alpha^n (B_n - B) \right| \leq$$

$$\lim_{\alpha\uparrow 1} (1-\alpha) \sum_{n=M}^{\infty} \alpha^n |B_n - B| \leq$$

$$\lim_{\alpha\uparrow 1} (1-\alpha) \sum_{n=M}^{\infty} \alpha^n \cdot \epsilon = \epsilon \cdot \lim_{\alpha\uparrow 1} \alpha^M = \epsilon.$$

(iii) Neem $B_n = (-1)^n,\ n \in \mathbf{N}_0$. Deze rij is niet convergent. Maar $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} B_k = 0$, dus de rij is wel Cesaro-convergent. Verder is

$$\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n B_n = \lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}(-\alpha)^n = \lim_{\alpha\uparrow 1}(1-\alpha)\frac{1}{1+\alpha} = 0.$$

Dus de rij is ook Abel-convergent.
(iv) Neem $\{B_n\}_{n=0}^{\infty} = 0, 1, -1, 2, -2, 3, -3, \ldots$, ofwel

$$\begin{cases} B_{2k-1} = k, & k \in \mathbf{N} \\ B_{2n} = -n, & n \in \mathbf{N}_0 \end{cases}$$

Nu is

$$\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n B_n = \lim_{\alpha\uparrow 1}(1-\alpha)\left[\sum_{k=1}^{\infty}\alpha^{2k-1}B_{2k-1} + \sum_{n=1}^{\infty}\alpha^{2n}B_{2n}\right] =$$

$$\lim_{\alpha\uparrow 1}(1-\alpha)\left[\sum_{k=1}^{\infty}\alpha^{2k-1}\cdot k - \sum_{n=1}^{\infty}\alpha^{2n}\cdot n\right] = \lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=1}^{\infty}\alpha^{2n-1}(1-\alpha)n =$$

$$\lim_{\alpha\uparrow 1}\alpha(1-\alpha)^2\sum_{n=1}^{\infty}\alpha^{2n}\cdot n.$$

Omdat

$$\sum_{n=0}^{\infty} n\alpha^n = \sum_{n=0}^{\infty}(n+1)\alpha^n - \sum_{n=0}^{\infty}\alpha^n = \sum_{n=0}^{\infty}\frac{\mathrm{d}}{\mathrm{d}\alpha}\alpha^{n+1} - \frac{1}{1-\alpha} = \frac{\mathrm{d}}{\mathrm{d}\alpha}\sum_{n=0}^{\infty}\alpha^{n+1} - \frac{1}{1-\alpha} =$$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{\alpha}{1-\alpha}\right) - \frac{1}{1-\alpha} = \frac{1}{(1-\alpha)^2} - \frac{1}{1-\alpha},$$

is $\sum_{n=0}^{\infty}\alpha^{2n}\cdot n = \frac{1}{(1-\alpha^2)^2} - \frac{1}{1-\alpha^2}$, en dus is

$$\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n B_n = \lim_{\alpha\uparrow 1}\alpha(1-\alpha)^2\left(\frac{1}{(1-\alpha^2)^2} - \frac{1}{1-\alpha^2}\right) = \lim_{\alpha\uparrow 1}\alpha\left(\frac{1}{(1+\alpha)^2} - \frac{1-\alpha}{1+\alpha}\right) = \frac{1}{4}.$$

Dus $\{B_n\}_{n=0}^{\infty}$ is Abel-convergent.
Maar als n oneven is, dan is $\frac{1}{n}\sum_{k=0}^{n-1} B_k = 0$, en als n even is, dan is $\frac{1}{n}\sum_{k=0}^{n-1} B_k = \frac{1}{n}B_{n-1} = \frac{1}{n}\cdot\frac{n}{2} = \frac{1}{2}$. Dus $\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} B_k = 0$ bestaat niet, dus $\{B_n\}_{n=0}^{\infty}$ is niet Cesaro-convergent. ♢

Stelling 3.3 *Laat P een stochastische matrix zijn. Dan geldt:*
(i) $P^* := \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} P^k$ bestaat, m.a.w. $P^n \overset{(C)}{\to} P^*$;
(ii) $P^*P = PP^* = P^*P^* = P^*$.

Bewijs
Zie [Kemeny en Snell, 1960][4]. ♢

P^* wordt de *stationaire matrix* van P genoemd.

4. J.G. Kemeny en J.L. Snell, *Finite Markov chains*, Van Nostrand, Princeton, 1960

Lemma 3.2 $\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n(P^n-P^*)=0$.

Bewijs
Er geldt dat

$$\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n P^* = P^*\lim_{\alpha\uparrow 1}(1-\alpha)\frac{1}{1-\alpha}=P^*.$$

Dus we moeten bewijzen dat

$$\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n P^n = P^*.$$

Uit stelling 3.3 volgt dat de rij $\{P^n\}_{n=0}^{\infty}$ Cesaro-convergent is met Cearo-limiet P^*, dus is, volgens stelling 3.1, $\{P^n\}_{n=0}^{\infty}$ ook Abel-convergent met Abel-limiet P^*, dus

$$\lim_{\alpha\uparrow 1}(1-\alpha)\sum_{n=0}^{\infty}\alpha^n P^n = P^*.$$

♢

Stelling 3.4 *Laat P een stochastische matrix zijn. Dan geldt:*
(i) $I-P+P^*$ is regulier en $Z:=(I-P+P^*)^{-1}$ voldoet aan

$$Z=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sum_{i=0}^{k-1}(P-P^*)^i.$$

(ii) $D:=Z-P^*$ voldoet aan

$$D=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sum_{i=0}^{k-1}(P^i-P^*)$$

en

$$P^*D=DP^*=(I-P)D+P^*-I=D(I-P)+P^*-I=0.$$

Bewijs
Zie [Kemeny en Snell, 1960]. ♢

Z en D worden de *fundamentele matrix* resp. *afwijkingsmatrix* genoemd. De naam 'afwijkingsmatrix' wordt duidelijk als we aannemen dat P irreducibel en aperiodiek is. In dat geval heeft P^* identieke rijen, en kan bewezen worden dat $D=\sum_{n=0}^{\infty}(P^n-P^*)$: zie [Veinott, 1974][5].

Lemma 3.3 *(i)* $Z=\lim_{\alpha\uparrow 1}\sum_{n=0}^{\infty}\alpha^n(P-P^*)^n$.
(ii) $D=\lim_{\alpha\uparrow 1}\sum_{n=0}^{\infty}\alpha^n(P^n-P^*)$.

5. A.F. Veinott Jr., Markov decision chains; G.B. Dantzig en B.C Eaves (eds),Studies in optimization, *Studies in Mathematics, deel 10: The Mathematical Association of America*, p. 124-159, 1974

Bewijs
(i) $\lim_{\alpha\uparrow 1}[I-\alpha(P-P^*)]^{-1} = (I-P+P^*)^{-1} = Z$.
(ii) $\lim_{\alpha\uparrow 1}\sum_{n=0}^{\infty}\alpha^n(P^n-P^*) = \sum_{n=0}^{\infty}(P^n-P^*) = D$. ♢

Tenslotte geldt de volgende relatie tussen de gemiddelde verwachte opbrengst $\phi(\pi^\infty)$, de totale verwachte verdisconteerde opbrengst $v^\alpha(\pi^\infty)$ en de totale verwachte opbrengst over T periodes $v^T(\pi^\infty)$ voor een stationaire strategie π^∞:

Stelling 3.5 *(i)* $\phi(\pi^\infty) = P^*(\pi)r(\pi)$;
(ii) $\phi(\pi^\infty) = \lim_{\alpha\uparrow 1}(1-\alpha)v^\alpha(\pi^\infty)$;
(iii) $v^T(\pi^\infty) = T\phi(\pi) + D(\pi)r(\pi) - P^T(\pi)D(\pi)r(\pi)$.

Bewijs
Het bewijs staat in [Kallenberg 2]. ♢

3.2 De Laurentreeksontwikkeling

We laten zien dat er een deterministische strategie f_0^∞ bestaat zdd. $v^\alpha(f) = v^\alpha\ \forall\alpha\in[\alpha_0,1)$ voor zekere α_0. f_0 is dan een *Blackwell-optimale* strategie. Er geldt:

Stelling 3.6 *Er zijn getallen* $\alpha_m,\alpha_{m-1},\ldots,\alpha_0,\alpha_{-1}$ *en deterministische strategieën* $f_m^\infty, f_{m-1}^\infty,\ldots,f_0^\infty$ *zdd.:*
(i) $0=\alpha_m<\alpha_{m-1}<\ldots<\alpha_0<\alpha_{-1}=1$;
(ii) $v^\alpha(f_j^\infty) = v^\alpha\ \forall\alpha\in[\alpha_j,\alpha_{j-1}),\ j=m,m-1,\ldots,0.$

Bewijs
Het bewijs staat in [Kallenberg 2]. Er wordt eerst aangetoond dat er een Blackwell-optimale strategie f_0^∞ bestaat. Vervolgens wordt m.b.v. de Heine-Borel-Lebesgue-overdekkingstheorie [Zaanen, 1967][6] een eindige open overdekking van [0,1] geconstrueerd. Hiermee wordt (ii) bewezen. ♢

Stelling 3.7 *Definieer* $u^{-1}(\pi) = P^*(\pi)r(\pi),\ u^0(\pi) = D(\pi)r(\pi)$, *en* $u^{k+1}(\pi) = -D(\pi)u^k(\pi),\ k\geq 0$. *Dan geldt voor* $\alpha_0(\pi)<\alpha<1$ *dat*

$$v^\alpha(\pi^\infty) = \frac{1}{\alpha}\sum_{k=-1}^{\infty}\left(\frac{1-\alpha}{\alpha}\right)^k u^k(\pi),\ \textit{waarbij}\ \alpha_0(\pi) = \frac{\|D(\pi)\|}{1+\|D(\pi)\|}.$$

Bewijs
Laat $x(\pi) := \frac{1}{\alpha}\sum_{k=-1}^{\infty}\left(\frac{1-\alpha}{\alpha}\right)^k u^k(\pi)$. Dan is

$$x(\pi) = \frac{\phi(\pi^\infty)}{1-\alpha} + \frac{D(\pi)}{\alpha}\sum_{k=0}^{\infty}\left[\frac{\alpha-1}{\alpha}D(\pi)\right]^k r(\pi)$$

voor $\|\frac{\alpha-1}{\alpha}D(\pi)\|<1$, d.w.z. $\frac{\|D(\pi)\|}{1+\|D(\pi)\|}<\alpha<1$.
Omdat $v^\alpha(\pi^\infty)$ de unieke oplossing is van het lineaire stelsel $[I-\alpha P(\pi)]x = r(\pi)$, is het voldoende om te laten zien dat $r(\pi)-[I-\alpha P(\pi)]x(\pi) = 0$. Dit wordt in [Kallenberg 2] aangetoond. ♢

6. A.C. Zaanen, *Integration*, North Holland, Amsterdam, 1967

Gevolg 3.1 $v^\alpha(\pi^\infty) = \frac{\phi(\pi^\infty)}{1-\alpha} + u^0(\pi) + \epsilon(\alpha)$, *waarbij* $\lim_{\alpha\uparrow 1} \epsilon(\alpha) = 0$.

3.3 Extra gevoelige criteria

Bij het gemiddelde-opbrengst-criterium (§1.1.3) wordt geen rekening gehouden met opbrengsten in een eindig aantal periodes. Zo wordt geen verschil gezien tussen bijv. de opbrengsten $0, 0, 0, 0, \ldots$ en $10, 10, 0, 0, 0, \ldots$. Daarom zijn er criteria nodig die onderscheid maken tussen zulke rijen opbrengsten: *extra gevoelige criteria*. Eén manier om een criterium te maken, is het gebruik van verdiscontering met $\alpha \uparrow 1$. Een andere manier is het gebruik van extra gevoelige gemiddelde-opbrengst-criteria. Hieronder worden zes criteria besproken. Het blijkt dat voor alle criteria, behalve nummer 5, deterministische optimale strategieën bestaan.

1. Bias-optimaliteit
Een strategie R_* wordt *bias-optimaal* genoemd als

$$\lim_{\alpha\uparrow 1}[v^\alpha(R_*) - v^\alpha] = 0.$$

Er kan bewezen worden dat bias-optimaliteit gemiddelde optimaliteit impliceert[7].

2. Blackwell-optimaliteit
R_* heet *Blackwell-optimaal* als $v^\alpha(R_*) = v^\alpha\ \forall\alpha$ voldoende dicht bij 1, m.a.w.

$$\exists\alpha_0 \in (0,1):\ v^\alpha(R_*) = v^\alpha\ \forall\alpha \in [\alpha_0, 1).$$

Uit Blackwell-optimaliteit volgt dus bias-optimaliteit. Het omgekeerde is niet waar: in [Kallenberg 2] staat een tegenvoorbeeld dat is overgenomen uit [Blackwell, 1962]. In [Blackwell, 1962] wordt ook bewezen dat er altijd een Blackwell-optimale strategie bestaat.

3. n-verdisconteerde optimaliteit
Criteria 1 en 2 zijn speciale gevallen van n-verdisconteerde optimaliteit, $n = -1, 0, 1, \ldots$. R_* heet *n-verdisconteerd optimaal* als

$$\lim_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*) - v^\alpha] = 0.$$

Invullen van $n = 0$ geeft dus het criterium van bias-optimaliteit. In [Veinott, 1969][8] wordt bewezen dat (-1)-verdisconteerde optimaliteit equivalent is met gemiddelde optimaliteit, en dat Blackwell-optimaliteit hetzelfde is als n-verdisconteerde optimaliteit voor alle $n \geq N := \#S$.

Lemma 3.4 *(i) n-verdisconteerde optimaliteit impliceert $(n-1)$-verdisconteerde optimaliteit;*
(ii) Blackwell-optimaliteit impliceert n-verdisconteerde optimaliteit voor alle n;
(iii) n-verdisconteerde optimaliteit is equivalent met het criterium

$$\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*) - v^\alpha(R)] \geq 0\ \forall R \in C,$$

7. D. Blackwell, Discrete dynamic programming, *Annals of Mathematical Statistics* **33**, p. 719-726, 1962
8. A.F. Veinott Jr., Discrete dynamic programming with sensitive discount optimality criteria, *Annals of Mathematical Statistics* **40**, p. 1635-1660, 1969

waarbij C de verzameling van alle strategieën is.

Bewijs
(i) Als R_* n-verdisconteerd optimaal is, dan is

$$\lim_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=0 \Rightarrow$$

$$\lim_{\alpha\uparrow 1}(1-\alpha)\cdot(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=0 \Rightarrow$$

$$\lim_{\alpha\uparrow 1}(1-\alpha)^{-n+1}[v^\alpha(R_*)-v^\alpha]=0,$$

en dus is R_* $(n-1)$-verdisconteerd optimaal.
(ii) Als R_* Blackwell-optimaal is, dan is

$$v^\alpha(R_*)=v^\alpha\ \forall\alpha\in[\alpha_0,1)\Rightarrow(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=0\ \forall\alpha\in[\alpha_0,1)\Rightarrow$$

$$\lim_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=\lim_{\alpha\uparrow 1}0=0,$$

en dus is R_* n-verdisconteerd optimaal.
(iii) $\underline{\Rightarrow:}$ Omdat $v^\alpha(R)\le v^\alpha\ \forall R$, is $v^\alpha(R_*)-v^\alpha(R)\ge v^\alpha(R_*)-v^\alpha$, en dus is

$$\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha(R)]\ge\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=$$

$$\lim_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=0.$$

$\underline{\Leftarrow:}$ Omdat $\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha(R)]\ge 0\ \forall R$, kunnen we voor R een Blackwell-optimale strategie f_0^∞ invullen waarvoor geldt dat $v^\alpha(f_0^\infty)=v^\alpha$ voor α voldoende dicht bij 1. Hieruit volgt dat

$$\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]\ge 0.$$

Omdat $v^\alpha(R_*)\le v^\alpha$, volgt hieruit dat

$$\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]\le 0,$$

en dus is

$$\lim_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(R_*)-v^\alpha]=0.$$

♢

4. Gemiddelde overtaking (inhalende) optimaliteit
R_* heet *gemiddeld overtaking optimaal* als

$$\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[v^t(R_*)-v^t(R)]\ge 0\ \forall R\in C,$$

waarbij $v^t(R)$ de totale verwachte opbrengst over t periodes is. In [Lippman, 1969][9] wordt bewezen dat gemiddelde overtaking optimaliteit equivalent is met bias-optimaliteit, en dus ook met 0-verdisconteerde optimaliteit.

9. S.A. Lippman, Criterion equivalence in discrete dynamic programming, *Operations Research* **17**, p. 920-923, 1969

5. Overtaking optimaliteit
R_* heet *overtaking optimaal* als

$$\liminf_{T\to\infty} \sum_{t=1}^{T} [v^t(R_*) - v^t(R)] \geq 0 \ \forall R \in C.$$

Er bestaat in het algemeen geen overtaking optimale strategie, wat wordt aangetoond in [Brown, 1965][10]. Verder geldt dat uit overtaking optimaliteit gemiddelde overtaking optimaliteit volgt: als R_* overtaking optimaal is, dan is

$$\liminf_{T\to\infty} \sum_{t=1}^{T} [v^t(R_*) - v^t(R)] \geq 0 \ \forall R \in C.$$

Definieer $A_T := \sum_{t=1}^{T} [v^t(R_*) - v^t(R)]$. Dus $\liminf_{T\to\infty} A_T \geq 0$, dus

$$\forall \epsilon > 0 \ \exists T_\epsilon : \ A_T \geq -\epsilon \ \forall T \geq T_\epsilon.$$

Omdat $\left|\frac{1}{T} A_T\right| < |A_T|$, volgt hieruit dat

$$\frac{1}{T} A_T \geq -\epsilon \ \forall T \geq T_\epsilon \Rightarrow$$

$$\forall \epsilon > 0 \ \exists T_\epsilon : \ \frac{1}{T} A_T \geq -\epsilon \ \forall T \geq T_\epsilon.$$

Dus $\liminf_{T\to\infty} \frac{1}{T} A_T \geq 0$, ofwel

$$\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [v^t(R_*) - v^t(R)] \geq 0 \ \forall R \in C,$$

en dus is R_* gemiddeld overtaking optimaal.

6. n-gemiddelde optimaliteit
Dit is een uitbreiding van gemiddelde overtaking optimaliteit. Definieer voor willekeurige strategie R, $t \in \mathbf{N}$ en $n = -1, 0, 1, \ldots$ de vector

$$v^{n,t}(R) := \begin{cases} v^t(R) & \text{voor } n = -1 \\ \sum_{s=1}^{t} v^{n-1,s}(R) & \text{voor } n = 0, 1, \ldots \end{cases}$$

R_* heet *n-gemiddeld optimaal* als

$$\liminf_{T\to\infty} \frac{1}{T} [v^{n,T}(R_*) - v^{n,T}(R)] \geq 0$$

voor alle $R \in C$ en $n = -1, 0, 1, \ldots$.
Uit deze definitie volgt dat (-1)-gemiddelde optimaliteit hetzelfde is als gemiddelde optimaliteit, en 0-gemiddelde optimaliteit hetzelfde als gemiddelde overtaking optimaliteit.
In [Sladky, 1974][11] wordt bewezen dat n-gemiddelde optimaliteit equivalent is met n-verdisconteerde optimaliteit.

---

10. B.W. Brown, On the iterative method of dynamic programming on a finite space discrete Markov process, *Annals of Methematical Statistics* **36**, p. 1279-1285, 1965
11. K. Sladky, On the set of optimal controls for Markov chains with rewards, *Kybernatika* **10**, p. 350-367, 1974

3.4 n-verdisconteerde optimaliteit en strategieverbetering

We noteren de verzameling deterministische strategieën met $C(D)$.
Verder gebruiken we een *lexicografische ordening*: $u(\rho)$ is lexicografisch niet-negatief (positief) als de eerste niet-nul-vector van $(u^{-1}, u^0, \ldots)$ niet-negatief (positief) is, d.w.z.:

$$\begin{cases} u(\rho) \geq_l 0 \text{ als } \liminf_{\rho \downarrow 0} \rho^{-k} u(\rho) \geq 0 \text{ voor } k = -1, 0, 1, \ldots \\ u(\rho) >_l 0 \text{ als } u(\rho) \geq_l 0 \text{ en } u(\rho) \neq 0. \end{cases}$$

Hieronder volgt een algoritme m.b.v. strategieverbetering om voor $n = -1, 0, 1, \ldots$ een deterministische, n-verdisconteerd optimale strategie te vinden, m.a.w. een strategie die over $C(D)$ de vector $(u^{-1}(f), u^0(f), \ldots, u^{n+1}(f))$ lexicografisch maximaliseert. Voor $n = -1$ levert dit een gemiddeld optimale strategie op, en voor $n = 0$ een bias-optimale strategie. Verder blijkt een n-verdisconteerd optimale strategie voor alle $n \geq N - 1$ een Blackwell-optimale strategie te zijn. Het algoritme staat in [Kallenberg 2], en is gemaakt door [Miller en Veinott, 1969][12].
We definiëren $y = \max_{S \times A}[r + Px]$ als volgt: $y_i = \max_{a \in A(i)}[r_{ia} + \sum_j p_{iaj} x_j]$. Verder definiëren we $\operatorname{argmax}_{S \times A}[r + Px] = \{f \mid \max_{S \times A}[r + Px] = r(f) + P(f)x\}$.

Algoritme voor een n-verdisconteerd optimale strategie m.b.v. strategieverbetering

1. Neem een willekeurige $f^\infty \in C(D)$.
2. Bereken $(u^{-1}(f), u^0(f), \ldots, u^{n+1}(f))$ als unieke oplossing van het lineaire stelsel

$$\begin{cases} [I - P(f)]x^{-1} &= 0 \\ x^{-1} + [I - P(f)]x^0 &= r(f) \\ x^{k-1} + [I - P(f)]x^k &= 0,\ 1 \leq k \leq n+1;\ P^*(f)x^{n+1} = 0 \end{cases} \tag{3.4}$$

3. (a) Als $\max_{S \times A}[Pu^{-1}(f) - u^{-1}(f)] > 0$:

$$A^{(-1)} = \operatorname{argmax}_{S \times A}[Pu^{-1}(f) - u^{-1}(f)];$$

   kies g uit $A^{(-1)}$ en ga naar stap 5.
   (b) Als $\max_{S \times A^{(-1)}}[r + Pu^0(f) - u^0(f) - u^{-1}(f)] > 0$:

$$A^{(0)} = \operatorname{argmax}_{S \times A^{(-1)}}[r + Pu^0(f) - u^0(f) - u^{-1}(f)];$$

   kies g uit $A^{(0)}$. $A := A^{(-1)}$ en ga naar stap 5.
   (c) Doe voor $k = 0, \ldots, n$:
   Als $\max_{S \times A^{(k)}}[Pu^{k+1}(f) - u^{k+1}(f) - u^k(f)] > 0$:

$$A^{(k+1)} = \operatorname{argmax}_{S \times A^{(k)}}[Pu^{k+1}(f) - u^{k+1}(f) - u^k(f)];$$

   Kies g uit $A^{(k+1)}$. $A := A^{(k)}$ en ga naar stap 5.
4. f^∞ is n-verdisconteerd optimaal (STOP).
5. $f(i) := g(i),\ i \in S$, en ga naar stap 2.

12. B.L. Miller en A.F. Veinott Jr., Discrete dynamic programming with a small interest rate, *Annals of Mathematical Statistics* **40**, p. 366-370, 1969

De correctheid van het algoritme volgt uit onderstaande lemma's en stellingen, die bewezen worden in [Kallenberg 2].

Stelling 3.8 *Het stelsel (3.4) heeft de unieke oplossing* $(u^{-1}(f), u^0(f), \ldots, u^{n+1}(f))$.

Voor de volgende lemma's en stellingen hebben we de volgende notaties nodig: definieer voor $f^\infty, g^\infty \in C(D)$:

$$\begin{aligned} \psi^{-1}(f,g) &= P(g)u^{-1}(f) - u^{-1}(f) \\ \psi^0(f,g) &= r(g) + P(g)u^0(f) - u^0(f) - u^{-1}(f) \\ \psi^k(f,g) &= P(g)u^k(f) - u^k(f) - u^{k-1}(f), k \geq 1 \end{aligned}$$

Lemma 3.5 *Voor iedere* $f^\infty, g^\infty \in C(D)$ *en iedere* $m \in \mathbf{N}$ *geldt:*

$$\alpha v^\alpha(g^\infty) = \sum_{k=-1}^{m-1} \rho^k \{u^k(f) + \sum_{t=1}^{\infty} \alpha^t P^{t-1}(g)\psi^k(f,g)\} + \rho^m \sum_{t=1}^{\infty} \alpha^t P^{t-1}(g)u^{m-1}(f).$$

Lemma 3.6 *Als* f^∞ *en* g^∞ *opeenvolgende strategieën in het algoritme zijn, dan is* $v^\rho(g^\infty) > v^\rho(f^\infty)$ *voor* ρ *klein genoeg.*

Lemma 3.7 *Als het algoritme eindigt met strategie* f^∞*, dan is* f^∞ *een* n*-verdisconteerd optimale strategie.*

Stelling 3.9 *Het algoritme stopt na een eindig aantal iteraties met een* n*-verdisconteerd optimale strategie.*

Bewijs:
Bij het bewijs worden lemma's 3.6 en 3.7 gebruikt. ♢

Lemma 3.8 *Als* $\psi^k(f,g) = 0$ *voor* $k = 1, 2, \ldots, N$*, dan is ook* $\psi^k(f,g) = 0$ *voor iedere* $k \geq N+1$.

Stelling 3.10 *Als het algoritme gebruikt wordt om een* $(N-1)$*-verdisconteerd optimale strategie* f^∞ *te bepalen, dan is* f^∞ *ook een Blackwell-optimale strategie.*

Bewijs:
Bij het bewijs worden lemma's 3.5, 3.6 en 3.8 gebruikt. ♢

Hoofdstuk 4

Geneste LP-problemen (het irreducibele geval)

4.1 Inleiding

In dit hoofdstuk behandelen we een methode om voor irreducibele Markovbeslissingsproblemen een Blackwell-optimale strategie te vinden m.b.v. geneste LP-problemen. Dit hoofdstuk is gebaserd op het artikel [Avrachenkov en Altman][1].

Aanname 4.1 *Voor willekeurige strategie π heeft de Markovketen geïnduceerd door overgangsmatrix $P(\pi)$ één recurrente klasse en er zijn geen transiënte toestanden, m.a.w. de Markovketen is irreducibel.*

We beschouwen het Markovbeslissingsprobleem (MBP) met de eindige toestandsruimte $S = \{1, \ldots, N\}$ en overgangskansen $p_{iaj} = \mathbf{P}\{X_{t+1} = j \mid X_t = i,\ Y_t = a\}$. Acties a kunnen gekozen worden uit de actieverzameling $A(i)$. Als in toestand i actie a wordt gekozen, is de directe opbrengst r_{ia}. Het is voldoende om alleen C_s, de klasse van stationaire strategieën, te bekijken: zie [Derman, 1970][2] en [Puterman, 1994][3].
Voor iedere $\pi \in C_s$ kunnen we definiëren: $\pi_{ia} := \mathbf{P}\{Y_t = a \mid X_t = i\}$.
Vaak gebruiken we alleen deterministische strategieën $f \in C_d$, die gedefinieerd kunnen worden als $f(i) = a,\ a \in A(i)$.
Voor willekeurige strategie $\pi \in C_s$ kunnen we de overgangsmatrix $P(\pi)$ en de opbrengstvector $r(\pi)$ definiëren door

$$p_{ij}(\pi) := \sum_{a \in A(i)} p_{iaj}\pi_{ia}$$

$$r_i(\pi) := \sum_{a \in A(i)} r_{ia}\pi_{ia}$$

De verwachte gemiddelde opbrengst, als gestart wordt in toestand i, is

$$g_i(\pi) := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} [P^{t-1}(\pi) r(\pi)]_i$$

1. K.E. Avrachenkov en E. Altman, Sensitive discount optimality via nested linear programs for ergodic Markov decision processes, *IDC'99 Proceedings (Adelaide, Australië)*, p. 53-58, 1999
2. C. Derman, *Finite state Markovian decision processes*, Academic Press, New York, 1970
3. M.L. Puterman, *Markov decision processes*, John Wiley & Sons, New York, 1994

en de verwachte verdisconteerde opbrengst is

$$v_i^{\alpha}(\pi) := \sum_{t=1}^{\infty} \alpha^{t-1}[P^{t-1}(\pi)r(\pi)]_i,$$

waarbij i de begintoestand is, en $\alpha \in (0,1)$ de verdisconteringsfactor.
Voor de rente ρ geldt weer: $\rho = \frac{1-\alpha}{\alpha}$. Er geldt dat

$$v^{\rho}(\pi) = (1+\rho)[\rho^{-1}u^{-1}(\pi) + \sum_{n=0}^{\infty} \rho^n u^n(\pi)], \tag{4.1}$$

met $u^{-1}(\pi) = P^*(\pi)r(\pi)$, $u^0(\pi) = D(\pi)r(\pi)$ en $u^n(\pi) = (-1)^n D(\pi)^{n+1}r(\pi)$ (zie ook Stelling 3.7).
De stationaire matrix $P^*(\pi)$ en afwijkingsmatrix $D(\pi)$ zijn gedefinieerd als:

$$P^*(\pi) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} P^{t-1}(\pi)$$

$$D(\pi) := (I - P(\pi) + P^*(\pi))^{-1} - P^*(\pi).$$

We gebruiken de volgende eigenschappen van $P^*(\pi)$ en $D(\pi)$ (zie ook Stelling 3.4):

$$P^*(\pi)D(\pi) = D(\pi)P^*(\pi) = 0; \tag{4.2}$$

$$(I - P(\pi))D(\pi) = D(\pi)(I - P(\pi)) = I - P^*(\pi). \tag{4.3}$$

Er geldt dat $u^{-1}(\pi) = g(\pi)$, en $u^0(\pi) = D(\pi)r(\pi)$ noemen we de *verwachte bias-opbrengstvector*.
De stationaire strategie π_* is *verdisconteerd optimaal* voor vaste $\alpha \in (0,1)$ als

$$v_i^{\alpha}(\pi_*) \geq v_i^{\alpha}(\pi) \ \forall i \in S \ \forall \pi \in C_s;$$

π_* is *gemiddeld optimaal* als

$$g_i(\pi_*) \geq g_i(\pi) \ \forall i \in S \ \forall \pi \in C_s.$$

Met behulp van de Laurentreeksontwikkeling kunnen we extra gevoelige optimaliteitscriteria definiëren:
Een strategie $\pi^* \in C_s$ is *n-verdisconteerd optimaal* voor zekere $n \in \mathbf{N}_0$ als

$[u^n(\pi^*)]_i \geq [u^n(\pi)]_i \ \forall i \in S \ \forall \pi$ zdd. π $(n-1)$-verdisconteerd optimaal is, waarbij

(-1)-verdisconteerde optimaliteit gedefinieerd is als gemiddelde optimaliteit.

Lemma 4.1 *Bovenstaande definitie is equivalent met de definitie van n-verdisconteerde optimaliteit uit §3.3.*

Bewijs
We kunnen $v^{\alpha}(\pi^{\infty})$ schrijven als

$$v^{\alpha}(\pi^{\infty}) = \frac{1}{\alpha} \sum_{k=-1}^{\infty} \left(\frac{1-\alpha}{\alpha}\right)^k u^k(\pi) = (1+\rho) \sum_{k=-1}^{\infty} \rho^k u^k(\pi).$$

Volgens Lemma 3.4 is π_* n-verdisconteerd optimaal $\Leftrightarrow$

$$\liminf_{\alpha\uparrow 1}(1-\alpha)^{-n}[v^\alpha(\pi_*)-v^\alpha(\pi)]\geq 0\ \forall\pi.$$

Dit is equivalent met

$$\liminf_{\rho\downarrow 0}\frac{1}{\rho^n}\left\{(1+\rho)\sum_{k=-1}^{\infty}\rho^k[u^k(\pi_*)-u^k(\pi)]\right\}\geq 0\ \forall\pi\ \Leftrightarrow$$

$$\liminf_{\rho\downarrow 0}\sum_{k=-1}^{\infty}\rho^{k-n}[u^k(\pi_*)-u^k(\pi)]\geq 0\ \forall\pi\ \Leftrightarrow$$

$$\liminf_{\rho\downarrow 0}\sum_{k=-1}^{n}\rho^{k-n}[u^k(\pi_*)-u^k(\pi)]\geq 0\ \forall\pi.$$

We passen nu inductie naar n toe.
Voor $n=-1$ geldt: π_* is (-1)-verdisconteerd optimaal $\Leftrightarrow$

$$\liminf_{\rho\downarrow 0}\ [u^{-1}(\pi_*)-u^{-1}(\pi)]\geq 0\Leftrightarrow u^{-1}(\pi_*)-u^{-1}(\pi)\geq 0\ \forall\pi.$$

Inductiestap: neem aan dat beide definities voor $(n-1)$-verdisconteerde optimaliteit equivalent zijn. Dan geldt:

$$\liminf_{\rho\downarrow 0}\sum_{k=-1}^{n-1}\rho^{k-n+1}[u^k(\pi_*)-u^k(\pi)]\geq 0\ \forall\pi\ \Leftrightarrow$$

$$\begin{cases} u^{-1}(\pi_*) & \geq & u^{-1}(\pi) & \forall\pi \\ u^0(\pi_*) & \geq & u^0(\pi) & \forall\pi \text{ met } u^{-1}(\pi_*)=u^{-1}(\pi) \\ & \vdots & & \\ u^{n-1}(\pi_*) & \geq & u^{n-1}(\pi) & \forall\pi \text{ met } u^{-1}(\pi_*)=u^{-1}(\pi),\ldots,u^{n-2}(\pi_*)=u^{n-2}(\pi) \end{cases}$$

Er geldt dat

$$\liminf_{\rho\downarrow 0}\sum_{k=-1}^{n}\rho^{k-n}[u^k(\pi_*)-u^k(\pi)]=\liminf_{\rho\downarrow 0}\sum_{k=-1}^{n-1}\rho^{k-n}[u^k(\pi_*)-u^k(\pi)]+u^n(\pi_*)-u^n(\pi).$$

Als

$$u^n(\pi_*)\geq u^n(\pi)\ \forall\pi \text{ met } u^{-1}(\pi_*)=u^{-1}(\pi),\ldots,u^{n-1}(\pi_*)=u^{n-1}(\pi),$$

dan volgt meteen dat

$$\liminf_{\rho\downarrow 0}\sum_{k=-1}^{n}\rho^{k-n}[u^k(\pi_*)-u^k(\pi)]\geq 0.$$

Omgekeerd, als

$$\liminf_{\rho\downarrow 0}\sum_{k=-1}^{n-1}\rho^{k-n}[u^k(\pi_*)-u^k(\pi)]+u^n(\pi_*)-u^n(\pi)\geq 0,$$

dan kunnen we voor π een $(n-1)$-verdisconteerd optimale strategie invullen, zodat

$$\liminf_{\rho \downarrow 0} \sum_{k=-1}^{n-1} \rho^{k-n}[u^k(\pi_*) - u^k(\pi)] = 0,$$

en dus is

$$u^n(\pi_*) \geq u^n(\pi) \ \forall \pi \text{ met } u^{-1}(\pi_*) = u^{-1}(\pi), \ldots, u^{n-1}(\pi_*) = u^{n-1}(\pi).$$

♢

Een strategie π^* is *Blackwell-optimaal* als

$$\exists \rho_0 > 0 : \ \forall \rho \in (0, \rho_0] : \ v^\rho(\pi^*) \geq v^\rho(\pi) \ \forall \pi \in C_s.$$

Deze definitite is equivalent met de definitie in §3.3.

Stelling 4.1 *Stel dat π_* $(N-2)$-verdisconteerd optimaal is. Dan is π_* Blackwell-optimaal, en π_* is n-verdisconteerd optimaal voor alle $n \geq -1$.*

Bewijs
Zie [Lamond en Puterman, 1989][4] en [Veinott, 1974][5]. ♢

Uit Aanname 4.1 volgt dat $P^*(\pi)$ identieke rijen heeft, en dat $g(\pi) = P^*(\pi)r(\pi)$ identieke elementen heeft. Daarom wordt de notatie $g(\pi)$ zowel voor de vector als voor één component ervan gebruikt.

4.2 Gemiddelde optimaliteit

Om een gemiddeld optimale strategie te vinden, moeten we de volgende LP-problemen oplossen zie ook §1.1.3):

$$\text{LP: } \min \left\{ \ \tilde{g} \ \middle| \ \tilde{g} + \textstyle\sum_j (\delta_{ij} - p_{iaj})\tilde{h}_j \geq r_{ia}, \ i \in S, \ a \in A(i) \ \right\} \tag{4.4}$$

$$\text{DLP: } \max \left\{ \sum_{i,a} r_{ia} x_{ia} \ \middle| \ \begin{array}{l} \sum_{i,a} (\delta_{ij} - p_{iaj}) x_{ia} = 0, \ j \in S \\ \sum_{i,a} x_{ia} = 1, \ x_{ia} \geq 0, \ i \in S, \ a \in A(i) \end{array} \right\} \tag{4.5}$$

Een optimale oplossingen van (4.4) noteren we met (g_*, h_*), en een optimale basisoplossing van (4.5) noteren we met x_*. g_* is gelijk aan de waardevector (zie Stelling 1.6). De deterministische, gemiddeld optimale strategie f kan als volgt bepaald worden:

$$f(i) = a_i, \text{ zdd. } x_{*ia_i} > 0, \ i \in S. \tag{4.6}$$

4. B.F. Lamond en M.L. Puterman, Generalised inverses in discrete time Markov decision processes, *SIAM J. Matrix Anal. Appl.*, **10**, p. 118-134, 1989
5. A.F. Veinott Jr., Markov decision chains, *Studies in optimization*, eds. G.B. Dantzig en B.C. Eaves, p. 124-159, 1974

4.3 Bias-optimaliteit

Hieronder volgt een methode om een bias-optimale strategie te berekenen, gebaseerd op het systeem van twee geneste LP-problemen. De methode komt uit [Kallenberg, 1983][6].
We lossen eerst (4.4) en (4.5) op om een stationaire, gemiddeld optimale strategie te berekenen. Definieer nu

$$A_0(i) := \{\ a \in A(i) \mid g_* + \textstyle\sum_j (\delta_{ij} - p_{iaj}) h_{*j} = r_{ia}\ \},\ i \in S. \tag{4.7}$$

We kunnen $A_0(i)$ berekenen door iedere (i,a) af te gaan. Verder geldt: als x_* een optimale basisoplossing van (4.5) is, dan is $a \in A_0(i)$ als $x_{*ia} > 0$. Hierover gaat de volgende propositie, die afgeleid is van Lemma 5.3.1 (ii) en (iii) uit [Kallenberg, 1983]:

Propositie 4.1 *Laat f een deterministische, stationaire, gemiddeld optimale strategie zijn. Dan geldt:*
(i) $f(i) \in A_0(i),\ i \in S$, en
(ii) de bijbehorende bias-vector kan berekend worden m.b.v. de formule $u^0(f) = h_* - P^*(f)h_*$.

Bewijs
(i) Volgens de beperkingen van (4.4) is $g_* + [I - P(f)]h_* \geq r(f)$. Stel dat $g_* + [I - P(f)]h_* > r(f)$ en f is gemiddeld optimaal. Uit Aanname 4.1 volgt dat $P^*(f)$ strikt positief is, dus als we aan de linkerkant vermenigvuldigen met $P^*(f)$, dan krijgen we

$$P^*(f)g_* + P^*(f)[I - P(f)]h_* > P^*(f)r(f) \Rightarrow g_* > \phi(f).$$

Tegenspraak. Dus $g_* + [I - P(f)]h_* = r(f)$, m.a.w. $f(i) \in A_0(i),\ i \in S$.
(ii) Uit deel (i) volgt dat

$$D(f)\{g_* + (I - P(f))h_*\} = D(f)r(f) = u^0(f).$$

Omdat $g_* = P^*(f)r(f)$, volgt hieruit dat

$$u^0(f) = D(f)P^*(f)r(f) + D(f)(I - P(f))h_*.$$

Uit (4.2) en (4.3) volgt nu dat

$$u^0(f) = (I - P^*(f))h_* = h_* - P^*(f)h_*.$$

♢

Uit de definitie van n-verdisconteerde optimaliteit volgt dat de stationaire strategie d bias-optimaal is (ofwel 0-verdisconteerd optimaal) als $u_i^0(d) = \max_f\{u_i^0(f) \mid g(f) = g_*\}$. Dus de gemiddeld optimale strategie die de optimale waarde geeft bij de bias-vector, is bias-optimaal. Maximalisatie van $u^0(f)$ is equivalent met maximalisatie van $-P^*(f)h_*$ over alle gemiddeld optimale strategieën f.

Stelling 4.2 *f is gemiddeld optimaal $\Leftrightarrow$ $f(i) \in A_0(i)\ \forall i \in S$.*

---

6. L.C.M. Kallenberg, *Linear programming and finite Markovian control problems*, Mathematical Centre Tracts 148, Amsterdam, 1983

Bewijs
$\Rightarrow$: zie Propositie 4.1.
$\Leftarrow$: Als $f(i) \in A_0(i)\ \forall i \in S$, dan geldt dat

$$g_* + [I - P(f)]h_* = r(f).$$

Van links vermenigvuldigen met $P^*(f)$ geeft:

$$g_* = P^*(f)r(f) = \phi(f),$$

en dus is f gemiddeld optimaal. $\diamondsuit$

Voor het bepalen van de optimale strategie moeten we volgens (4.6) $f(i) = a_i$ kiezen zdd. $x_{*ia_i} > 0$. Volgens de orthogonaliteitsrelaties moet dan dus gelden dat

$$g_* + \sum\nolimits_j (\delta_{ij} - p_{iaj})h_{*j} = r_{ia}.$$

We kunnen ons dus beperken tot de acties a waarvoor $a \in A_0(i)$. Nu kunnen we een ander MBP beschouwen met toestandsruimte S en de verkleinde actieverzamelingen $A_0(i)$, $i \in S$. Uit propositie 4.1 volgt dat $A_0(i) \neq \emptyset\ \forall i \in S$. Alle strategieën die uit $A_0(i)$ worden geconstrueerd, induceren ook een irreducibele Markovketen. Het nieuwe verkleinde MBP is dus ook irreducibel, en we kunnen de LP-problemen (4.4) en (4.5) weer gebruiken:

$$\text{LP: } \min\left\{ \tilde{g}^{(0)} \;\middle|\; \tilde{g}^{(0)} + \textstyle\sum_j (\delta_{ij} - p_{iaj})\tilde{h}_j^{(0)} \geq -h_{*i},\ i \in S,\ a \in A_0(i) \right\}. \tag{4.8}$$

De optimale oplossing $g_*^{(0)}$ hiervan is de optimale gemiddelde opbrengst voor het verkleinde model.

$$\text{DLP: } \max\left\{ \sum_i (-h_{*i}) \sum_{a \in A_0(i)} x_{ia}^{(0)} \;\middle|\; \begin{array}{l} \sum_i \sum_{a \in A_0(i)} (\delta_{ij} - p_{iaj}) x_{ia}^{(0)} = 0,\ j \in S \\ \sum_i \sum_{a \in A_0(i)} x_{ia}^{(0)} = 1 \\ x_{ia}^{(0)} \geq 0,\ i \in S,\ a \in A_0(i) \end{array} \right\}. \tag{4.9}$$

Als $x_*^{(0)}$ een optimale basisoplossing van (4.9) is, dan is een deterministische strategie f, zdd. $f_*(i) = a_i$ als $x_{*ia_i}^{(0)} > 0$, een gemiddeld optimale strategie voor het verkleinde model. Uit Propositie 4.1 en Stelling 4.2 volgt nu:

Stelling 4.3 *Laat f_* een gemiddeld optimale strategie zijn voor het verkleinde MBP. Dan is $u^0(f_*) = u_{opt}^0$, waar u_{opt}^0 de optimale waarde van de bias-vector is. M.a.w., f_* is bias-optimaal.*

4.4 n-verdisconteerde optimaliteit

In deze paragraaf volgt een algoritme voor het bepalen van een n-verdisconteerd optimale strategie, gebaseerd op geneste LP-problemen. Het is een generalisatie van de methode uit paragraaf 3. De methode is iteratief: het LP-probleem voor de berekening van een $(n+1)$-verdisconteerd optimale strategie wordt geconstrueerd uit de oplossing van het LP-probleem voor een n-verdisconteerd optimale strategie. Het LP-probleem voor een n-verdisconteerd optimale strategie is:

$$\min\left\{ \tilde{g}^{(n)} \;\middle|\; \tilde{g}^{(n)} + \textstyle\sum_j (\delta_{ij} - p_{iaj})\tilde{h}_j^{(n)} \geq -h_{*i}^{(n-1)},\ i \in S,\ a \in A_n(i) \right\} \tag{4.10}$$

en het duale probleem is:

$$\max\left\{\sum_i(-h_{*i}^{(n-1)})\sum_{a\in A_n(i)}x_{ia}^{(n)}\;\middle|\;\begin{array}{l}\sum_i\sum_{a\in A_n(i)}(\delta_{ij}-p_{iaj})x_{ia}^{(n)}=0,\ j\in S\\ \sum_i\sum_{a\in A_n(i)}x_{ia}^{(n)}=1,\\ x_{ia}^{(n)}\geq 0,\ i\in S,\ a\in A_n(i)\end{array}\right\} \tag{4.11}$$

Als we in (4.10) en (4.11) $n=0$ invullen, en $h_*^{(-1)}:=h_*$ definiëren, dan krijgen we de vergelijkingen (4.8) en (4.9) terug.
Voor de inductiestap definiëren we nu eerst een nieuwe verkleinde actieverzameling: laat $(g_*^{(n)},h_*^{(n)})$ een optimale oplossing van (4.10) zijn. Dan kunnen we definiëren:

$$A_{n+1}(i):=\left\{\ a\in A_n(i)\ \middle|\ g_*^{(n)}+\textstyle\sum_j(\delta_{ij}-p_{iaj})h_{*j}^{(n)}=-h_{*i}^{(n-1)}\ \right\},\ i\in S \tag{4.12}$$

De inductiestap is gebaseerd op het volgende lemma, dat een generalisatie is van propositie 4.1:

Lemma 4.2 *Laat f een deterministische, stationaire, n-verdisconteerd optimale strategie zijn. Dan geldt:*
(i) $f(i)\in A_{n+1}(i),\ i\in S$, en
(ii) de bijbehorende $(n+1)$-verdisconteerde vector $u^{n+1}(f)$ kan berekend worden m.b.v. de formule

$$u^{n+1}(f)=[I-P^*(f)]h_*^{(n)} \tag{4.13}$$

Bewijs
(i) Volgens de beperkingen van (4.10) is $g_*^{(n)}+[I-P(f)]h_*^{(n)}\geq -h_{*i}^{(n-1)}$. Stel dat $g_*^{(n)}+[I-P(f)]h_*^{(n)}>-h_{*i}^{(n-1)}$ en f is n-verdisconteerd optimaal. Uit Aanname 4.1 volgt dat $P^*(f)$ strikt positief is, dus als we aan de linkerkant vermenigvuldigen met $P^*(f)$, dan krijgen we

$$P^*(f)g_*^{(n)}+P^*(f)[I-P(f)]h_*^{(n)}>-P^*(f)h_{*i}^{(n-1)}\Rightarrow g_*^{(n)}>-P^*(f)h_{*i}^{(n-1)}.$$

Tegenspraak. Dus $g_*^{(n)}+[I-P(f)]h_*^{(n)}=-h_{*i}^{(n-1)}$, m.a.w. $f(i)\in A_{n+1}(i),\ i\in S$.
(ii) Uit (i) volgt voor willekeurige n-verdisconteerd optimale strategie dat

$$g_*^{(n)}+(I-P(f))h_*^{(n)}=-h_*^{(n-1)} \tag{4.14}$$

Het bewijs gaat met inductie. Volgens Propositie 4.1(ii) klopt (4.13) voor $n=-1$. Stel dat (4.13) klopt voor $n=k-1$. Er geldt dat

$$u^{k+1}(f)=(-1)^{k+1}D^{k+2}(f)r(f)=-D(f)(-1)^kD^{k+1}(f)r(f)=-D(f)u^k(f).$$

Volgens de inductieveronderstelling is $u^k(f)=[I-P^*(f)]h_*^{(k-1)}$. M.b.v. (4.14), (4.2) en (4.3) en de eigenschap $g_*^{(k)}=P^*(f)h_*^{(k-1)}$ kunnen we nu schrijven:

$$u^{k+1}(f)=-D(f)u^k(f)=-D(f)[I-P^*(f)]h_*^{(k-1)}=D(f)(-h_*^{(k-1)})=$$

$$D(f)[g_*^{(k)}+(I-P(f))h_*^{(k)}]=D(f)P^*(f)h_*^{(k-1)}+D(f)(I-P(f))h_*^{(k)}=$$

$$(I-P^*(f))h_*^{(k)}.$$

Dus (4.13) klopt ook voor $n = k$. $\diamondsuit$

Uit de definitie van $(n+1)$-verdisconteerde optimaliteit volgt: als de strategie f n-verdisconteerd optimaal is en $u^{n+1}(f)$ bereikt bovendien zijn maximale waarde, dan is f $(n+1)$-verdisconteerd optimaal. Volgens lemma 4.2 is de maximalisatie van $u^{n+1}(f)$ equivalent met de maximalisatie van $-P^*(f)h_*^{(n)}$.
$g^{(n+1)}(f) := P^*(f)(-h_*^{(n)})$ kan beschouwd worden als de vector van de verwachte gemiddelde opbrengst voor het MBP met overgangsmatrix $P(f)$, directe-opbrengst-vector $r_{ia}^{(n+1)} := -h_{*i}^{(n)}$ en de verkleinde actieverzameling $A_{n+1}(i)$. Volgens lemma 4.2(i) is $A_{n+1}(i) \neq \emptyset \ \forall i \in S$. Dit verkleinde MBP noemen we het $(n+1)$-staps verkleinde MBP.
Volgens aanname 4.1 is het n-staps verkleinde MBP irreducibel voor alle $n \geq -1$. De optimale waarde van de verwachte gemiddelde opbrengst kan bepaald worden m.b.v. het LP-probleem

$$\min \left\{ \ \tilde{g}^{(n+1)} \ \middle| \ \tilde{g}^{(n+1)} + \textstyle\sum_j (\delta_{ij} - p_{iaj})\tilde{h}_j^{(n+1)} \geq -h_{*i}^{(n)}, \ i \in S, \ a \in A_{n+1}(i) \ \right\} \tag{4.15}$$

en de gemiddeld optimale strategie voor het $(n+1)$-staps verkleinde MBP kan bepaald worden m.b.v. het duale probleem

$$\max \left\{ \sum_i (-h_{*i}^{(n)}) \sum_{a \in A_{n+1}(i)} x_{ia}^{(n+1)} \ \middle| \ \begin{array}{l} \sum_{i,a} (\delta_{ij} - p_{iaj}) x_{ia}^{(n+1)} = 0, \ j \in S \\ \sum_{i,a} x_{ia}^{(n+1)} = 1 \\ x_{ia}^{(n+1)} \geq 0, \ i \in S, \ a \in A_{n+1}(i) \end{array} \right\} : \tag{4.16}$$

als $x_*^{(n+1)}$ een optimale basisoplossing is van (4.16), dan is de deterministische strategie f_*, zdd.

$$f_*(i) = a_i \text{ als } x_{*ia_i}^{(n+1)} > 0, \tag{4.17}$$

gemiddeld optimaal voor het $(n+1)$-staps verkleinde MBP.
Volgens de volgende stelling is deze strategie ook $(n+1)$-optimaal voor het oorspronkelijke probleem:

Stelling 4.4 *Laat f een gemiddeld optimale strategie zijn voor het $(n+1)$-staps verkleinde MBP. Dan is*

$$u^{n+1}(f_*) = u_{opt}^{n+1},$$

waar u_{opt}^{n+1} de optimale waarde van de $(n+1)$-verdisconteerde vector is. M.a.w., f_* is $(n+1)$-verdisconteerd optimaal.

Bewijs
Laat d een stationaire, $(n+1)$-verdisconteerd optimale strategie zijn voor het oorspronkelijke MBP. Volgens o.a. [Puterman, 1994] en [Veinott, 1969][7] bestaat zo'n strategie. d is natuurlijk ook n-verdisconteerd optimaal. Uit Lemma 4.2(ii) volgt dat

$$u_{\text{opt}}^{n+1} = u^{n+1}(d) = h_*^{(n)} - P^*(d)h_*^{(n)}.$$

Omdat f_* een gemiddeld optimale strategie is voor het $(n+1)$-staps verkleinde MBP, geldt dat

$$u_{\text{opt}}^{n+1} \geq u^{n+1}(f_*) = h_*^{(n)} - P^*(f_*)h_*^{(n)} \geq h_*^{(n)} - P^*(d)h_*^{(n)}.$$

7. A.F. Veinott Jr., Discrete dynamic programming with sensitive discount optimality criteria, *Ann. Math. Stat.* **40**, p. 1635-1660, 1969

Hieruit volgt dat

$$u_{\text{opt}}^{n+1} = u^{n+1}(f_*).$$

♢

4.5 Algoritme voor de berekening van een Blackwell-optimale strategie

In deze paragraaf geven we een formeel algoritme voor de berekening van een Blackwell-optimale strategie. n-verdisconteerd optimale strategieën zijn bijproducten hiervan. Als we alleen een n-verdisconteerd optimale strategie willen berekenen, kunnen we volstaan met maximaal n iteraties.

Algoritme:

1. Los LP-probleem (4.5) op om de gemiddeld optimale strategie te bepalen. Als het een unieke optimale basisoplossing heeft: STOP. De bijbehorende strategie is Blackwell-optimaal en n-verdisconteerd optimaal $\forall n \geq -1$. Anders: ga naar stap 2.
2. Bereken de oplossing van (4.4).
3. Neem $n := 0$.
4. Bepaal m.b.v. (4.7) en (4.12) de verkleinde actieverzameling A_n.
5. Los LP-probleem (4.11) op voor het n-staps verkleinde MBP. Bepaal m.b.v. (4.17) de n-verdisconteerd optimale strategie.
   Als alleen een n-verdisconteerd optimale strategie wordt gezocht: STOP.
   Als de optimale basisoplossing uniek is, of als $n = \#S - 2$: STOP. In dit geval is een Blackwell-optimale strategie gevonden.
   Anders: ga naar stap 6.
6. Los LP-probleem (4.10) op voor het n-staps verkleinde MBP.
7. $n := n + 1$ en ga naar stap 4.

Uit stelling 4.1 volgt dat het algoritme eindig is. De stappen 4-7 worden maximaal $\#S - 2$ keer herhaald.

Hoofdstuk 5

Relatie tussen strategieverbetering en geneste LP-problemen

5.1 Gemiddelde optimaliteit

Deze paragraaf is gebaseerd op een deel van §4.6 uit het boek [Kallenberg, 1983]. We nemen aan dat voor alle $f^\infty \in C_D$ de Markovketen geïnduceerd door $P(f)$ irreducibel is.
Bekijk nu het algoritme m.b.v. strategieverbetering (SV) uit §3.4. Omdat $\phi(f^\infty)$ identieke componenten heeft, kunnen we $\phi(f^\infty)$ vervangen door $\phi_0(f^\infty) \cdot e$, met $\phi_0(f^\infty) \in \mathbf{R}$. We definiëren

$$A(i,f) := \{a \in A(i) \mid \phi_0(f^\infty) + \sum\nolimits_j (\delta_{ij} - p_{iaj}) u_j(f^\infty) < r_{ia}\} =$$

$$\{a \in A(i) \mid r_{ia} + \sum\nolimits_j p_{iaj} u_j(f^\infty) - u_i(f^\infty) - \phi_0(f^\infty) > 0\}.$$

In het algoritme in §3.4 wordt dit

$$\{a \in A(i) \mid r_{ia} + \sum\nolimits_j p_{iaj} u_j^0(f^\infty) - u_i^0(f^\infty) - u^{-1}(f^\infty) > 0\}.$$

Als $A(i,f) = \emptyset$, dan is $g(i) := f(i)$. Anders: kies $g(i)$ uit $A(i,f)$. Volgens Stelling 4.6.1 uit [Kallenberg, 1983] is $x(f)$ een extreme toelaatbare oplossing van het LP-probleem (4.5). Het duale probleem hiervan is

$$\min\{\tilde{\phi} \mid \tilde{\phi} + \sum\nolimits_j (\delta_{ij} - p_{iaj}) \tilde{u}_j \geq r_{ia}\},$$

zie ook (4.4). Omdat $x_{if(i)}(f) > 0 \ \forall i \in S$, volgt uit de orthogonaliteitsrelaties dat

$$d_{if(i)} := \tilde{\phi} + \sum\nolimits_j (\delta_{ij} - p_{if(i)j}) \tilde{u}_j - r_{if(i)} = 0 \ \forall i \in S.$$

Hieruit volgt dat

$$\tilde{\phi} \cdot e = P^*(f)(\tilde{\phi} \cdot e) = P^*(f)[r(f) - (I - P(f))\tilde{u}] = P^*(f) r(f) = \phi(f^\infty),$$

en

$$\phi(f^\infty) + (I - P(f))u(f) = \phi(f^\infty) + (I - P(f))D(f)r(f) =$$
$$\phi(f^\infty) + (I - P^*(f))r(f) = r(f).$$

Dus $(I - P(f))(u(f^\infty) - \tilde{u}) = 0$, waaruit volgt dat

$$u(f^\infty) - \tilde{u} = P(f)(u(f^\infty) - \tilde{u}) \;\Rightarrow\; u(f^\infty) - \tilde{u} = P^*(f)(u(f^\infty) - \tilde{u}).$$

Omdat $P^*(f)$ identieke rijen heeft, heeft $u(f^\infty) - \tilde{u}$ identieke componenten, zodat

$$\sum\nolimits_j p_{iaj}(u(f^\infty) - \tilde{u}) = [u(f^\infty) - \tilde{u}]_i \;\Rightarrow\; \sum\nolimits_j (\delta_{ij} - p_{iaj})\tilde{u}_j = \sum\nolimits_j (\delta_{ij} - p_{iaj})u_j(f^\infty).$$

Nu is dus

$$d_{ia} = \phi_0(f^\infty) + \sum\nolimits_j (\delta_{ij} - p_{iaj})u_j(f^\infty) - r_{ia}.$$

Omdat $a \in A(i, f) \;\Leftrightarrow\; d_{ia} < 0$, correspondeert de verzameling acties waaruit $g(i)$ gekozen kan worden met de mogelijke keuzes voor de pivotkolom in de simplexmethode. Hieruit volgt:

Stelling 5.1 *Elk algoritme m.b.v. strategieverbetering is equivalent met een 'block-pivoting' simplexalgoritme, d.w.z. een simplexalgoritme waarbij meerdere pivots tegelijk kunnen worden gekozen.*

5.2 Een rekenvoorbeeld

Om te onderzoeken wat de relatie tussen beide algoritmes is als we een n-verdisconteerd optimale strategie zoeken, kunnen we een eenvoudig MBP bekijken waarbij iedere strategie gemiddeld optimaal is:

Voorbeeld
$S = \{1, 2, 3, 4\},\ A(i) = \{1, 2\}\ \forall i,\ p_{112} = p_{213} = p_{314} = p_{411} = 1,\ p_{122} = p_{123} = 1/2,\ p_{223} = 1/3,\ p_{224} = 2/3,\ p_{321} = 1/4,\ p_{324} = 3/4,\ p_{421} = 1/4,\ p_{422} = 3/4$ en de overige overgangskansen zijn 0.
Nu moeten we de directe opbrengsten r_{ia} zó kiezen, dat alle $2^4 = 16$ deterministische, stationaire strategieën gemiddeld optimaal zijn, m.a.w. $P^*(f)r(f) = 1\ \forall f$. Dit levert het volgende stelsel vergelijkingen op:

$$\begin{pmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 4 & 0 & 4 & 0 & 0 & 4 \\
4 & 0 & 4 & 0 & 0 & 4 & 3 & 0 \\
7 & 0 & 16 & 0 & 0 & 16 & 0 & 12 \\
3 & 0 & 0 & 3 & 1 & 0 & 3 & 0 \\
3 & 0 & 0 & 12 & 4 & 0 & 0 & 12 \\
12 & 0 & 0 & 12 & 0 & 4 & 11 & 0 \\
15 & 0 & 0 & 48 & 0 & 16 & 0 & 44 \\
0 & 2 & 1 & 0 & 2 & 0 & 2 & 0 \\
0 & 2 & 7 & 0 & 8 & 0 & 0 & 8 \\
0 & 4 & 2 & 0 & 0 & 4 & 3 & 0 \\
0 & 14 & 25 & 0 & 0 & 32 & 0 & 24 \\
0 & 6 & 0 & 3 & 4 & 0 & 6 & 0 \\
0 & 6 & 0 & 21 & 10 & 0 & 0 & 24 \\
0 & 6 & 0 & 3 & 0 & 4 & 5 & 0 \\
0 & 6 & 0 & 15 & 0 & 8 & 0 & 16
\end{pmatrix}
\begin{pmatrix}
r_{11} \\ r_{12} \\ r_{21} \\ r_{22} \\ r_{31} \\ r_{32} \\ r_{41} \\ r_{42}
\end{pmatrix}
=
\begin{pmatrix}
4 \\ 13 \\ 15 \\ 51 \\ 10 \\ 31 \\ 39 \\ 123 \\ 7 \\ 25 \\ 13 \\ 95 \\ 19 \\ 61 \\ 18 \\ 45
\end{pmatrix}$$

Dit stelsel heeft oneindig veel oplossingen, waaronder de oplossing $r_{11} = 1,\ r_{12} = 1/2,\ r_{21} = 0,\ r_{22} = -2/3,\ r_{31} = 0,\ r_{32} = 1/2,\ r_{41} = 3,\ r_{42} = 3$. Als we nu het algoritme uit §3.4 toepassen, met $n = 0$ en met beginstrategie (1,1,1,1) (de i-de component geeft aan welke actie gekozen wordt in toestand i), dan krijgen we:
3(a) Geen resultaat; $A^{(-1)} = A$.
3(b) Geen resultaat; $A^{(0)} = A$.
3(c) $(k = 0)$ $i = 1:\ \max\{0, -3/8\} = 0$
$i = 2:\ \max\{0, 1/6\} = 1/6$
$i = 3:\ \max\{0, 5/16\} = 5/16$
$i = 4:\ \max\{0, -9/16\} = 0$
Dus we nemen in de tweede iteratie strategie (1,2,2,1).

3(a) Geen resultaat; $A^{(-1)} = A$.
3(b) Geen resultaat; $A^{(0)} = A$.
3(c) $(k = 0)$ $i = 1:\ \max\{0, -1/3\} = 0$
$i = 2:\ \max\{0, 0\} = 0$
$i = 3:\ \max\{-1/3, 0\} = 0$
$i = 4:\ \max\{0, -1/2\} = 0$
Dus strategie (1,2,2,1) is bias-optimaal.

Start nu met strategie (1,2,2,2).
3(a) Geen resultaat; $A^{(-1)} = A$.
3(b) Geen resultaat; $A^{(0)} = A$.
3(c) $(k = 0)$ $i = 1:\ \max\{0, -0.3252\} = 0$
$i = 2:\ \max\{0.1951, 0\} = 0.1951$
$i = 3:\ \max\{-0.4472, 0\} = 0$
$i = 4:\ \max\{0.6341, 0\} = 0.6341$
Dus we nemen in de tweede iteratie strategie (1,1,2,1).

3(a) Geen resultaat; $A^{(-1)} = A$.
3(b) Geen resultaat; $A^{(0)} = A$.
3(c) $(k = 0)$ $i = 1:\ \max\{0, -1/3\} = 0$
$i = 2:\ \max\{0, 0\} = 0$
$i = 3:\ \max\{-1/3, 0\} = 0$
$i = 4:\ \max\{0, -1/2\} = 0$
Dus ook strategie (1,1,2,1) is bias-optimaal.

Omdat we bij toestand 2 geen verschil zien tussen actie 1 en 2, maar bij de overige toestanden wel, zijn alleen de strategieën (1,1,2,1) en (1,2,2,1) bias-optimaal.

Om een 1-verdisconteerd optimale strategie te vinden, nemen we nu $n = 1$, en starten met strategie (1,2,2,1). We hoeven alleen stap 3(c) met $k = 1$ en $i = 2$ te bekijken: $\max\{4/13, 0\} = 4/13$
Dus we nemen in de volgende iteratie strategie (1,1,2,1):
$\max\{0, -4/15\} = 0$.
Dus alleen strategie (1,1,2,1) is 1-verdisconteerd optimaal, en dus ook Blackwell-optimaal.

Nu gaan we het algoritme uit §4.5 toepassen:

1. Probleem (4.5) wordt: maximaliseer

$$x_{11} + \frac{1}{2}x_{12} - \frac{2}{3}x_{22} + \frac{1}{2}x_{32} + 3x_{41} + 3x_{42}$$

onder de voorwaarden

$$\left\{\begin{array}{rrrrrrrrcl}
x_{11} & +x_{12} & & & & -\frac{1}{4}x_{32} & -x_{41} & -\frac{1}{4}x_{42} & = & 0 \\
-x_{11} & -\frac{1}{2}x_{12} & +x_{21} & +x_{22} & & & & -\frac{3}{4}x_{42} & = & 0 \\
 & -\frac{1}{2}x_{12} & -x_{21} & -\frac{1}{3}x_{22} & +x_{31} & +x_{32} & & & = & 0 \\
 & & & -\frac{2}{3}x_{22} & -x_{31} & -\frac{3}{4}x_{32} & +x_{41} & +x_{42} & = & 0 \\
x_{11} & +x_{12} & +x_{21} & +x_{22} & +x_{31} & +x_{32} & +x_{41} & +x_{42} & = & 1 \\
 & & & & & & & x_{ia} & \geq & 0 \ \forall i,a
\end{array}\right.$$

Omdat iedere deterministische, stationaire strategie optimaal is, kunnen we een basisoplossing kiezen: kies

$$x_{*12} = x_{*22} = x_{*32} = x_{*42} = 0 \ \Rightarrow x_{*11} = x_{*21} = x_{*31} = x_{*41} = 1/4.$$

2. Een optimale oplossing van (4.4) wordt nu:

$$h_{*1} = h_{*2} = 0,\ h_{*3} = 1,\ h_{*4} = 2,\ g_* = 1.$$

3. $n = 0$.
4. $A_0(i) = \{1, 2\}\ \forall i$.
5. Probleem (4.9) kunnen we schrijven als:
maximaliseer

$$-x_{31} - x_{32} - 2x_{41} - 2x_{42}$$

onder de voorwaarden

$$\left\{\begin{array}{rrrrrrrrrcl}
x_{11} & +x_{12} & & & & -\frac{1}{4}x_{32} & -x_{41} & -\frac{1}{4}x_{42} & +z_1 & = & 0 \\
-x_{11} & -\frac{1}{2}x_{12} & +x_{21} & +x_{22} & & & & -\frac{3}{4}x_{42} & +z_2 & = & 0 \\
 & -\frac{1}{2}x_{12} & -x_{21} & -\frac{1}{3}x_{22} & +x_{31} & +x_{32} & & & +z_3 & = & 0 \\
 & & & -\frac{2}{3}x_{22} & -x_{31} & -\frac{3}{4}x_{32} & +x_{41} & +x_{42} & +z_4 & = & 0 \\
x_{11} & +x_{12} & +x_{21} & +x_{22} & +x_{31} & +x_{32} & +x_{41} & +x_{42} & +z_5 & = & 1 \\
 & & & & & & & & x_{ia} & \geq & 0 \ \forall i,a
\end{array}\right.$$

Omdat moet gelden dat $z_i = 0\ \forall i$, moeten we eerst $z_0 := -z_1 - z_2 - z_3 - z_4 - z_5$ maximaliseren. Het eerste simplextableau wordt nu (de pivot wordt met * aangegeven):

		x_{11}	x_{12}	x_{21}	x_{22}	x_{31}	x_{32}	x_{41}	x_{42}
z_1	0	1*	1	0	0	0	−1/4	−1	−1/4
z_2	0	−1	−1/2	1	1	0	0	0	−3/4
z_3	0	0	−1/2	−1	−1/3	1	1	0	0
z_4	0	0	0	0	−2/3	−1	−3/4	1	1
z_5	1	1	1	1	1	1	1	1	1
x_0	0	0	0	0	0	1	1	2	2
z_0	−1	−1	−1	−1	−1	−1	−1	−1	−1

→

		z_1	x_{12}	x_{21}	x_{22}	x_{31}	x_{32}	x_{41}	x_{42}
x_{11}	0	1	1	0	0	0	$-1/4$	-1	$-1/4$
z_2	0	1	$1/2$	1	1	0	$-1/4$	-1	-1
z_3	0	0	$-1/2$	-1	$-1/3$	1	1	0	0
z_4	0	0	0	0	$-2/3$	-1	$-3/4$	$1*$	1
z_5	1	-1	0	1	1	1	$5/4$	2	$5/4$
x_0	0	0	0	0	0	1	1	2	2
z_0	-1	1	0	-1	-1	-1	$-5/4$	-2	$-5/4$

$\rightarrow$

		z_1	x_{12}	x_{21}	x_{22}	x_{31}	x_{32}	z_4	x_{42}
x_{11}	0	1	1	0	$-2/3$	-1	-1	1	$3/4$
z_2	0	1	$1/2$	1	$1/3$	-1	-1	1	0
z_3	0	0	$-1/2$	-1	$-1/3$	$1*$	1	0	0
x_{41}	0	0	0	0	$-2/3$	-1	$-3/4$	1	1
z_5	1	-1	0	1	$7/3$	3	$11/4$	-2	$-3/4$
x_0	0	0	0	0	$4/3$	3	$5/2$	-2	0
z_0	-1	1	0	-1	$-7/3$	-3	$-11/4$	2	$3/4$

$\rightarrow$

		z_1	x_{12}	x_{21}	x_{22}	z_3	x_{32}	z_4	x_{42}
x_{11}	0	1	$1/2$	-1	-1	1	0	1	$3/4$
z_2	0	1	0	0	0	1	0	1	0
x_{31}	0	0	$-1/2$	-1	$-1/3$	1	1	0	0
x_{41}	0	0	$-1/2$	-1	-1	1	$1/4$	1	1
z_5	1	-1	$3/2$	$4*$	$10/3$	-3	$-1/4$	-2	$-3/4$
x_0	0	0	$3/2$	3	$7/3$	-3	$-1/2$	-2	0
z_0	-1	1	$-3/2$	-4	$-10/3$	3	$1/4$	2	$3/4$

$\rightarrow$

De volgende tableaus horen bij strategie (1,1,1,1) resp. (1,1,2,1); de rij van z_2 kunnen we nu weglaten:

		z_1	x_{12}	z_5	x_{22}	z_3	x_{32}	z_4	x_{42}
x_{11}	$1/4$	$3/4$	$7/8$	$1/4$	$-1/6$	$1/4$	$-1/16$	$1/2$	$9/16$
x_{31}	$1/4$	$-1/4$	$-1/8$	$1/4$	$1/2$	$1/4$	$15/16*$	$-1/2$	$-3/16$
x_{41}	$1/4$	$-1/4$	$-1/8$	$1/4$	$-1/6$	$1/4$	$3/16$	$1/2$	$13/16$
x_{21}	$1/4$	$-1/4$	$3/8$	$1/4$	$5/6$	$-3/4$	$-1/16$	$-1/2$	$-3/16$
x_0	$-3/4$	$3/4$	$3/8$	$-3/4$	$-1/6$	$-3/4$	$-5/16$	$-1/2$	$9/16$
z_0	0	0	0	1	0	0	0	0	0

$\rightarrow$

		z_1	x_{12}	z_5	x_{22}	z_3	x_{31}	z_4	x_{42}
x_{11}	$4/15$	$11/15$	$13/15$	$4/15$	$-2/15$	$4/15$	$1/15$	$7/15$	$11/20$
x_{32}	$4/15$	$-4/15$	$-2/15$	$4/15$	$8/15$	$4/15$	$16/15$	$-8/15$	$-1/5$
x_{41}	$1/5$	$-1/5$	$-1/10$	$1/5$	$-4/15$	$1/5$	$-1/5$	$3/5$	$17/20$
x_{21}	$4/15$	$-4/15$	$11/30$	$4/15$	$13/15*$	$-11/15$	$1/15$	$-8/15$	$-1/5$
x_0	$-2/3$	$2/3$	$1/3$	$-2/3$	0	$-2/3$	$1/3$	$-2/3$	$1/2$
z_0	0	0	0	1	0	0	0	0	0

$\rightarrow$

Het volgende tableau hoort bij strategie (1,2,2,1); we kunnen de rij van z_0 nu ook weglaten:

		z_1	x_{12}	z_5	x_{21}	z_3	x_{31}	z_4	x_{42}
x_{11}	4/13	9/13	12/13	4/13	2/13	2/13	1/13	5/13	27/52
x_{32}	4/39	−4/39	−14/39	4/39	−8/13	28/39	40/39	−8/39	−1/13
x_{41}	11/39	−11/39	1/78	11/39	4/13	−1/39	−7/39	17/39	41/52
x_{22}	4/13	−4/13	11/26	4/13	15/13	−11/13	1/13	−8/13	−3/13
x_0	−2/3	2/3	1/3	−2/3	0	−2/3	1/3	−2/3	1/2

Als we nu bij de laatste drie tableaus naar de rij van x_0 kijken, dan valt op dat de getallen onder de niet-basisvariabelen x_{ia} precies tegengesteld zijn aan de getallen die bij het algoritme uit §3.4 bij dezelfde strategie in stap 3(c) ($k = 0$) berekend zijn. Ook nu blijkt zowel strategie (1,1,2,1) als strategie (1,2,2,1) bias-optimaal te zijn.

6. Een optimale oplossing van (4.8) is af te lezen uit het laatste simplextableau:

$$h_{*1}^{(0)} = 2/3,\ h_{*2}^{(0)} = 0,\ h_{*3}^{(0)} = -2/3,\ h_{*4}^{(0)} = -2/3,\ g_*^{(0)} = -2/3.$$

7. $n = 1$.

Volgende iteratie:

4. $A_1(1) = A_1(4) = \{1\},\ A_1(3) = \{2\},\ A_1(2) = \{1, 2\}$.

5. Probleem (4.11) wordt nu:

$$\text{maximaliseer } -\frac{2}{3}x_{11} - 0(x_{21} + x_{22}) + \frac{2}{3}x_{32} + \frac{2}{3}x_{41}$$

onder de voorwaarden

$$\begin{cases} x_{11} & & & -\frac{1}{4}x_{32} & -x_{41} & +z_1 & = & 0 \\ -x_{11} & +x_{21} & +x_{22} & & & +z_2 & = & 0 \\ & -x_{21} & -\frac{1}{3}x_{22} & +x_{32} & & +z_3 & = & 0 \\ & & -\frac{2}{3}x_{22} & -\frac{3}{4}x_{32} & +x_{41} & +z_4 & = & 0 \\ x_{11} & +x_{21} & +x_{22} & +x_{32} & +x_{41} & +z_5 & = & 1 \\ & & & & & x_{ia} & \geq & 0\ \forall i\ \forall a \end{cases}$$

Na wat rekenwerk krijgen we de tableaus die corresponderen met strategie (1,1,2,1) resp. (1,2,2,1):

		z_1	z_2	x_{22}	z_5	z_4
x_{11}	4/15	7/15	−4/15	−2/15	4/15	1/5
x_{21}	4/15	7/15	11/15	13/15	4/15	1/5
z_3	0	1	1	0	0	1
x_{41}	1/5	−2/5	−1/5	−4/15	1/5	2/5
x_{32}	4/15	−8/15	−4/15	8/15	4/15	−4/5
x_0	2/15	−14/15	−2/15	4/15	2/15	−2/5

		z_1	z_2	x_{21}	z_5	z_4
x_{11}	4/13	7/13	−2/13	2/13	4/13	3/13
x_{22}	4/13	7/13	11/13	15/13	4/13	3/13
z_3	0	1	1	0	0	1
x_{41}	11/39	−10/39	1/39	4/13	11/39	6/13
x_{32}	4/39	−32/39	−28/39	−8/13	4/39	−12/13
x_0	2/39	−14/13	−14/39	−4/13	2/39	−6/13

Ook nu is het getal onder de niet-basisvariabele x_{ia} precies tegengesteld aan het getal dat bij het algoritme uit §3.4 bij dezelfde strategie in stap 3(c) ($k = 1$) berekend is. Alleen strategie (1,1,2,1) is 1-verdisconteerd optimaal, dus ook Blackwell-optimaal. Dit voorbeeld geeft het vermoeden dat beide strategieën ook bij n-verdisconteerde optimaliteit equivalent werken, $n \geq 0$. Dit wordt bewezen in de volgende paragraaf.

5.3 n-verdisconteerde optimaliteit

We bekijken nu het algemene geval van n-verdisconteerde optimaliteit. Als we $n = 0$ nemen, krijgen we bias-optimaliteit.
Bekijk weer het algoritme m.b.v. strategieverbetering (SV) uit §3.4. We definiëren nu

$$A_n(i,f) := \{a \in A(i) \mid \sum\nolimits_j p_{iaj} u_j^{n+1}(f) - u_i^{n+1}(f) - u_i^n(f) > 0.\}$$

Als $A_n(i,f) = \emptyset$, dan is $g(i) := f(i)$. Anders: kies $g(i)$ uit $A_n(i,f)$. Een strategie f in het SV-algoritme correspondeert met een oplossing x van (4.11) waarvoor geldt dat $x_{if(i)} > 0 \; \forall i \in S$ en $x_{ia} = 0 \; \forall i \in S \; \forall a \neq f(i)$. Bekijk nu het simplextableau dat correspondeert met strategie f. Het getal dat in de kolom van een niet-basisvariabele x_{ia}, en in de rij van de doelfunctie staat, heeft de waarde

$$d_{ia}^{(n)} := \tilde{g}^{(n)} + \sum\nolimits_j (\delta_{ij} - p_{iaj}) \tilde{h}_j^{(n)} + h_{*i}^{(n-1)}.$$

Hierbij is $h_*^{(-1)} := h_*$. Omdat $x_{if(i)}(f) > 0 \; \forall i \in S$, volgt uit de orthogonaliteitsrelaties dat

$$d_{if(i)}^{(n)} = 0 \; \forall i \in S.$$

Omdat $\tilde{g}^{(n)}$ de gemiddelde opbrengst is bij strategie f, geldt nu dat

$$\tilde{g}^{(n)} \cdot e = P^*(f)\tilde{g}^{(n)} \cdot e = -P^*(f)[h_*^{(n-1)} - (I - P(f))\tilde{h}^{(n)}] = -P^*(f)h_*^{(n-1)} = g_*^{(n)} \cdot e,$$

en (gebruik Propositie 4.1(ii) en Lemma 4.2(ii))

$$g_*^{(n)} \cdot e + (I - P(f))u^{n+1}(f) = g_*^{(n)} \cdot e - (I - P(f))D(f)u^n(f) =$$

$$g_*^{(n)} \cdot e - (I - P(f))D(f)(I - P^*(f))h_*^{(n-1)} =$$

$$g_*^{(n)} \cdot e - (I - P^*(f))^2 h_*^{(n-1)} = g_*^{(n)} \cdot e - (I - P^*(f))h_*^{(n-1)} = -h_*^{(n-1)}.$$

Dus $(I - P(f))(u^{n+1}(f) - \tilde{h}^{(n)}) = 0$, waaruit volgt dat

$$u^{n+1}(f) - \tilde{h}^{(n)} = P(f)(u^{n+1}(f) - \tilde{h}^{(n)}) \;\Rightarrow\; u^{n+1}(f) - \tilde{h}^{(n)} = P^*(f)(u^{n+1}(f) - \tilde{h}^{(n)}).$$

Omdat $P^*(f)$ identieke rijen heeft, heeft $u^{n+1}(f) - \tilde{h}^{(n)}$ identieke componenten, zodat

$$\sum\nolimits_j (\delta_{ij} - p_{iaj}) \tilde{h}_j^{(n)} = \sum\nolimits_j (\delta_{ij} - p_{iaj}) u_j^{n+1}(f).$$

Nu is dus

$$d_{ia}^{(n)} = \tilde{g}^{(n)} + \sum\nolimits_j (\delta_{ij} - p_{iaj}) u_j^{n+1}(f) + h_{*i}^{(n-1)}.$$

Volgens Propositie 4.1(ii) en Lemma 4.2(ii) geldt voor een deterministische, stationaire, $(n-1)$-verdisconteerd optimale strategie f dat

$$u^n(f) = [I - P^*(f)]h_*^{(n-1)}.$$

Bovendien is hierboven voor $n \geq 0$ bewezen dat

$$\tilde{g}^{(n)} \cdot e = -P^*(f)h_*^{(n-1)}.$$

Nu is $d_{ia}^{(n)}$ te schrijven als

$$d_{ia}^{(n)} = [(I - P^*(f))h_*^{(n-1)}]_i + \sum\nolimits_j (\delta_{ij} - p_{iaj})u_j^{n+1}(f) =$$

$$\sum\nolimits_j (\delta_{ij} - p_{iaj})u_j^{n+1}(f) + u_i^n(f).$$

Omdat $a \in A_n(i, f) \Leftrightarrow d_{ia}^{(n)} < 0$, correspondeert de verzameling acties waaruit $g(i)$ gekozen kan worden met de mogelijke keuzes voor de pivotkolom in de simplexmethode. Hieruit volgt:

Stelling 5.2 *Het algoritme m.b.v. strategieverbetering uit §3.4 is equivalent met het algoritme m.b.v. geneste LP-problemen uit §4.5.*

Hoofdstuk 6

Literatuurlijst

K.E. Avrachenkov en E. Altman, Sensitive discount optimality via nested linear programs for ergodic Markov decision processes, *IDC'99 Proceedings (Adelaide, Australië)*, p. 53-58, 1999

D. Bertsimas en J.N. Tsitsiklis, *Introduction to linear programming*, Athena Scientific, 1997

D. Blackwell, Discrete dynamic programming, *Annals of Mathematical Statistics* **33**, p. 719-726, 1962

D. Blackwell, Discrete Dynamic programming, *Annals of Mathematical Statistics* **36**, p. 226-235, 1965

B.W. Brown, On the iterative method of dynamic programming on a finite space discrete Markov process, *Annals of Methematical Statistics* **36**, p. 1279-1285, 1965

C. Derman, *Finite state Markovian decision processes*, Academic Press, New York, 1970

C.C. Gonzaga, Path-following methods for linear programming, *SIAM-review* **34**, p. 167-224, 1992

L.C.M. Kallenberg, *Stochastische Dynamische Programmering*, Univ. Leiden, 1982/ 1983

L.C.M. Kallenberg, *Linear programming and finite Markovian control problems*, Mathematical Centre Tracts 148, Amsterdam, 1983

L.C.M. Kallenberg, *Inleiding Besliskunde*, Univ. Leiden, 2001

L.C.M. Kallenberg, *Markov Decision Theory*, Univ. Leiden

J.G. Kemeny en J.L. Snell, *Finite Markov chains*, Van Nostrand, Princeton, 1960

B.F. Lamond en M.L. Puterman, Generalised inverses in discrete time Markov de-

cision processes, *SIAM J. Matrix Anal. Appl.*, **10**, p. 118-134, 1989

S.A. Lippman, Criterion equivalence in discrete dynamic programming, *Operations Research* **17**, p. 920-923, 1969

B.L. Miller en A.F. Veinott Jr., Discrete dynamic programming with a small interest rate, *Annals of Mathematical Statistics* **40**, p. 366-370, 1969

S. Mizuno, M.J. Todd en Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, *Mathematics of Operations Research* **18**, p. 964-981, 1993

R.E. Powell en S.M. Shah, *Summability theory and applications*, Van Nostrand Reinhold, Londen, 1972

M.L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York, 1994

K. Sladky, On the set of optimal controls for Markov chains with rewards, *Kybernatika* **10**, p. 350-367, 1974

S. Vavasis en Y. Ye, A primal-dual interior-point method whose running time depends only on the constraint matrix, *Mathematical Programming* **74**, p. 79-120, 1996

A.F. Veinott Jr., Discrete dynamic programming with sensitive discount optimality criteria, *Annals of Mathematical Statistics* **40**, p. 1635-1660, 1969

A.F. Veinott Jr., Markov decision chains, *Studies in optimization*, eds. G.B. Dantzig en B.C. Eaves, p. 124-159, 1974

A.F. Veinott Jr., Markov decision chains; G.B. Dantzig en B.C Eaves (eds), Studies in optimization, *Studies in Mathematics, deel 10: The Mathematical Association of America*, p. 124-159, 1974

D. Widder, *Laplace transform*, Princeton University Press, Princeton, New Jersey, 1946

Y. Ye, *Interior Point Algorithms: Theory and Analysis*, JohnWiley & Sons, 1997

Y. Ye, A New Complexity Result on Solving the Markov Decision Problem, 2002 (niet gepubliceerd)

A.C. Zaanen, *Integration*, North Holland, Amsterdam, 1967