Supplementary Material: Composing Value Functions in Reinforcement Learning

Benjamin van Niekerk [*1] **Steven James** [*1] **Adam Earle** [1] **Benjamin Rosman** [1 2]

1. The Value Function

Lemma 3. *If π is a proper policy then V_π is bounded and $V_{\pi,n} \to V_\pi$ as $n \to \infty$.*

Proof. Since both the reward function and the KL-divergence term are bounded we can find an $M > 0$ such that $|r - \tau \text{KL}[\pi||\bar{\pi}]| < M$. Then, for any state s in $\mathcal{S}$ we have

$$\sum_{t=0}^{\infty} \int |r - \tau \text{KL}[\pi||\bar{\pi}]| dP_s^\pi \leq M \sum_{t=0}^{\infty} P_s^\pi(s_t \notin \mathcal{G}) < \infty. \tag{1}$$

Hence, by Rudin (1987, Theorem 1.38), $V_{\pi,n} \to V_\pi$ as $n \to \infty$. Finally, taking the supremum of (1) over $\mathcal{S}$ proves that V_π is bounded. □

Lemma 4. *Let π be a continuous proper policy and $f : \mathcal{S} \to \mathbb{R}$ be a bounded measurable function. Then*

$$\sup_{s \in \mathcal{S}} \left| \sum_{t=n+1}^{\infty} \mathbb{E}_s^\pi[f(s_t)] \right| \to 0 \text{ as } n \to \infty.$$

Proof. For each state s in $\mathcal{S}$ we have

$$|\mathbb{E}_s^\pi[f(s_t)]| \leq ||f||_\infty P_s^\pi(s_t \notin \mathcal{G}).$$

It follows that

$$\sup_{s \in \mathcal{S}} \left| \sum_{t=n+1}^{\infty} \mathbb{E}_s^\pi[f(s_t)] \right| \leq \sup_{s \in \mathcal{S}} \sum_{t=n+1}^{\infty} |\mathbb{E}_s^\pi[f(s_t)]| \leq ||f||_\infty \sup_{s \in \mathcal{S}} \sum_{t=n+1}^{\infty} P_s^\pi(s_t \notin \mathcal{G}).$$

Since π is proper the sum $\sum_{t=0}^{\infty} P_s^\pi(s_t \notin \mathcal{G})$ converges uniformly on $\mathcal{S}$ so the tail $\sum_{t=n+1}^{\infty} P_s^\pi(s_t \notin \mathcal{G})$ must converge uniformly to 0 as $n \to \infty$. □

Corollary 2. *Let π be a proper policy and $f : \mathcal{S} \to \mathbb{R}$ be a bounded measurable function. Then $\mathbb{E}_s^\pi[f(s_t)] \to 0$ as $t \to \infty$ for every state s in $\mathcal{S}$.*

2. The Bellman Operators

Lemma 5. *(Hernández-Lerma & Muñoz de Ozak, 1992, Remark 2.2) Let Assumption 1 hold and let $f : \mathcal{S} \to \mathbb{R}$ be an upper semicontinuous and bounded from above. Then the map*

$$a \mapsto \int_{\mathcal{S}} f(s') \rho_{(s,a)}(ds')$$

[*]Equal contribution [1]School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa [2]Council for Scientific and Industrial Research, Pretoria, South Africa. Correspondence to: Benjamin van Niekerk <benjamin.vanniekerk@students.wits.ac.za>, Steven James <steven.james@wits.ac.za>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

is upper semicontinuous for each state s in $\mathcal{S}$.

Lemma 6. *Let π be a policy and let $V_1, V_2, \ldots, V, U$ be bounded, real-valued and measurable functions on $\mathcal{S}$ then:*

(a) $\mathcal{T}_\pi$ is monotone, $U \leq V$ implies $\mathcal{T}_\pi U \leq \mathcal{T}_\pi V$.

(b) $\mathcal{T}$ is monotone, $U \leq V$ implies $\mathcal{T}U \leq \mathcal{T}V$.

(c) If $V_n \uparrow V$ then $\mathcal{T}V_n \uparrow \mathcal{T}V$.

(d) If $V_n \downarrow V$ then $\mathcal{T}V_n \downarrow \mathcal{T}V$

Proof. (a) Basic property of integrals. (b) Follows from (a).

(c) Suppose that $V_n \uparrow V$. Then for any policy π, an application of the monotone convergence theorem gives $\mathcal{T}_\pi V_n \uparrow \mathcal{T}_\pi V$. Making use of Hinderer (1970, Lemma 3.14) to interchange the limit and supremum we have

$$\begin{aligned}\lim_{n\to\infty}[\mathcal{T}V_n](s) &= \lim_{n\to\infty}\sup_\pi[\mathcal{T}_\pi V_n](s)\\ &= \sup_\pi \lim_{n\to\infty}[\mathcal{T}_\pi V_n](s)\\ &= \sup_\pi[\mathcal{T}_\pi V](s) = [\mathcal{T}V](s),\end{aligned}$$

for every state s in $\mathcal{S}$. Furthermore, by part (b), the operator $\mathcal{T}$ is monotone so $\mathcal{T}V_n \uparrow \mathcal{T}V$.

(d) Since $\log$ and $\exp$ are continuous functions the monotone convergence theorem gives us $\mathcal{L}V_n \downarrow \mathcal{L}V$. But, by Lemma 1, $\mathcal{T}V_n = \mathcal{L}V_n$ so $\mathcal{T}V_n \downarrow \mathcal{T}V$. □

Lemma 7. *Let π be a proper policy and $V : \mathcal{S} \to \mathbb{R}$ be a bounded measurable function. Then $\mathcal{T}_\pi^n V \to V_\pi$ as $n \to \infty$.*

Proof. For every state s it follows, from the definition of $\mathcal{T}_\pi$, that

$$[\mathcal{T}_\pi^n V](s) = \mathbb{E}^\pi_{s,n}\left[\sum_{t=0}^{n-1}\left(r(s_t,a_t) - \tau\mathrm{KL}[\pi_{s_t}||\bar{\pi}_{s_t}]\right) + V(s_n)\right].$$

Since π is proper, by Corollary 2, the expectation $\mathbb{E}^\pi_{s,n}[V(s_n)]$ converges to 0 as $n \to \infty$. Now, taking limits we have, by Lemma 3, that

$$\lim_{n\to\infty}[\mathcal{T}_\pi^n V](s) = \lim_{n\to\infty}\mathbb{E}^\pi_{s,n}\left[\sum_{t=0}^{n-1} r(s_t,a_t) - \tau\mathrm{KL}[\pi_{s_t}||\bar{\pi}_{s_t}]\right] = V_\pi(s).$$

Therefore, $\mathcal{T}_\pi^n V$ converges pointwise to V_π. □

Lemma 8. *Let π be a proper policy, then the family of functions $\{\mathcal{T}_\pi^n V_\pi\}_{n=1}^\infty$ is uniformly bounded.*

Proof. First note that $\mathcal{T}_\pi^n V_\pi$ is well-defined for all n since, by Lemma 3, V_π is bounded and measurable. By Corollary 2 we can find an N in $\mathbb{N}$ such that $|\mathbb{E}^\pi_{s,n}[V(s_n)]| < 1$ for all $n \geq N$. Put $K := \max_{n<N}|\mathbb{E}^\pi_{s,n}[V(s)]|$, then using Lemma 3 we have

$$\begin{aligned}|[\mathcal{T}_\pi^n V_\pi](s)| &= \left|\mathbb{E}^\pi_{s,n}\left[\sum_{t=0}^{n-1}\left(r(s_t,a_t) - \tau\mathrm{KL}[\pi_{s_t}||\bar{\pi}_{s_t}]\right) + V_\pi(s_n)\right]\right|\\ &\leq \mathbb{E}^\pi_{s,n}\left[\sum_{t=0}^{n-1}\left|r(s_t,a_t) - \tau\mathrm{KL}[\pi_{s_t}||\bar{\pi}_{s_t}]\right|\right] + \left|\mathbb{E}^\pi_{s,n}\left[V_\pi(s_n)\right]\right|\\ &\leq M\sum_{t=0}^{\infty} P_s^\pi(s_t \notin \mathcal{G}) + 1 + K\end{aligned}$$

for all states s. Taking the supremum over $\mathcal{S}$ and using the definition of proper policies proves the result. □

Lemma 9. *Let π be a proper policy. Then the value function V_π is the unique bounded, measurable solution of the Bellman equation $V_\pi = \mathcal{T}_\pi V_\pi$.*

Proof. First we show that V_π satisfies the Bellman equation. By the definition of $\mathcal{T}_\pi$ we have

$$[\mathcal{T}_\pi^{n+1} V_\pi](s) = \int_{\mathcal{A}} \left(r(s,a) + \int_{\mathcal{S}} [\mathcal{T}_\pi^n V_\pi](s') p_{(s,a)}(ds') \right) \pi_s(da) - \tau \mathrm{KL}[\pi_s || \bar{\pi}_s].$$

By Lemma 8 the family of functions $\{\mathcal{T}_\pi^n V_\pi\}_{n=1}^\infty$ is uniformly bounded. Using the bounded convergence theorem and Lemma 7 we have

$$V_\pi(s) = \int_{\mathcal{A}} \left(r(s,a) + \int_{\mathcal{S}} V_\pi(s') p_{(s,a)}(ds') \right) \pi_s(da) - \tau \mathrm{KL}[\pi_s || \bar{\pi}_s].$$

Therefore, $V_\pi = \mathcal{T}_\pi V_\pi$. For uniqueness, suppose that V is bounded and measure and that $V = \mathcal{T}_\pi V$. Then $V = \mathcal{T}_\pi^n V$ and letting $n \to \infty$ we get $V = V_\pi$ by Lemma 7. □

Lemma 10. *Let Assumption 2 hold. Let π be a stationary policy and $V : \mathcal{S} \to \mathbb{R}$ be a bounded measurable function such that $\mathcal{T}_\pi V \geq V$. Then π is proper and $V \leq V_\pi$.*

Proof. Since $\mathcal{T}_\pi$ is monotonic we have that

$$V(s) \leq [\mathcal{T}_\pi^n V](s) = \mathbb{E}_{s,n}^\pi \left[\sum_{t=0}^{n-1} \left(r(s_t, a_t) - \tau \mathrm{KL}[\pi_{s_t} || \bar{\pi}_{s_t}] \right) + V(s_n) \right]. \tag{2}$$

Taking the limit superior as $n \to \infty$ gives

$$\begin{aligned} V(s) &\leq \limsup_{n \to \infty} \mathbb{E}_{s,n}^\pi \left[\sum_{t=0}^{n-1} r(s_t, a_t) - \tau \mathrm{KL}[\pi_{s_t} || \bar{\pi}_{s_t}] \right] + ||V||_\infty \\ &= V_\pi(s) + ||V||_\infty. \end{aligned}$$

If π is improper then V_π is unbounded below, contradicting the above. Therefore π is proper. Finally, in view of Lemma 10, taking the limit as $n \to \infty$ in (2) gives $V \leq V_\pi$. □

Lemma 11. *If $V : \mathcal{S} \to \mathbb{R}$ is a bounded measurable function satisfying the optimality equation $V = \mathcal{T} V$ then V is unique.*

Proof. Suppose that there are bounded measurable functions V and V' satisfying $V = \mathcal{T} V$ and $V' = \mathcal{T} V'$ respectively. Then by Lemma 1 there are policies $\pi = \mathcal{B}[V]$ and $\pi' = \mathcal{B}[V']$ such that $V = \mathcal{T}_\pi V$ and $V' = \mathcal{T}_{\pi'} V'$. By Lemma 10 both π and π' are proper so, by Lemma 9, $V = V_\pi$ and $V' = V_{\pi'}$. But, by definition, $V = \mathcal{T}^n V \geq \mathcal{T}_{\pi'}^n V$ for all n in $\mathbb{N}$. Taking $n \to \infty$ gives $V \geq V'$. The same argument shows that $V' \geq V$ so $\mathcal{T}$ has at most one fixed point. □

Lemma 12. *Let Assumptions 1 and 2 hold and suppose that the optimal value function is bounded above. If $\{V_{\pi_n}\}_{n=1}^\infty$ is a sequence of value functions generated from soft-policy iteration starting from any proper policy, then $V_{\pi_n} \uparrow V$ and V is the unique bounded fixed point of $\mathcal{T}$.*

Proof. Suppose that π_0 is a given proper policy π_0. We use π_0 as the initial policy for a policy iteration procedure. Suppose, that we have generated n proper policies $\pi_0, \ldots, \pi_{n-1}$ via soft-policy iteration with

$$V_{\pi_0} \leq \mathcal{T} V_{\pi_0} \leq V_{\pi_1} \leq \ldots \leq V_{\pi_{n-2}} \leq \mathcal{T} V_{\pi_{n-2}} \leq V_{\pi_{n-1}}. \tag{3}$$

Then we choose $\pi_n = \mathcal{B}[V_{\pi_{n-1}}]$. Then, by Lemma 1, the policy π_n satisfies $\mathcal{T}_{\pi_n} V_{\pi_{n-1}} = \mathcal{T} V_{\pi_{n-1}}$ and we have that

$$V_{\pi_{n-1}} = \mathcal{T}_{\pi_{n-1}} V_{\pi_{n-1}} \leq \mathcal{T} V_{\pi_{n-1}} = \mathcal{T}_{\pi_n} V_{\pi_{n-1}}.$$

It follows, by Lemma 10, that π_n is proper. Also, by the monotonicity of $\mathcal{T}_{\pi_n}$ we have

$$V_{\pi_n} = \lim_{m \to \infty} \mathcal{T}_{\pi_n}^m V_{\pi_{n-1}} \geq \mathcal{T}_{\pi_n} V_{\pi_{n-1}} = \mathcal{T} V_{\pi_{n-1}} \geq V_{\pi_{n-1}}.$$

Therefore, we can use soft policy iteration to construct a sequence $\{\pi_n\}_{n=1}^{\infty}$ of proper policies with

$$V_{\pi_n} \leq \mathcal{T}V_{\pi} \leq V_{\pi_{n+1}}. \tag{4}$$

Since the sequence of value functions is non-decreasing V_{π_n} converges to some function V with $V \leq V^*$. By assumption V^* is bounded above so V is bounded. From Lemma 6 we have $\mathcal{T}V_{\pi_n} \uparrow \mathcal{T}V$. Taking the limit of (4) as $n \to \infty$ shows that $V = \mathcal{T}V$. □

Lemma 13. *Let Assumptions 1 and 2 hold and suppose that the optimal value function is bounded above. If $\{V_{\pi^n}\}_{n=1}^{\infty}$ is generated by soft-policy iteration starting from any proper policy then there exists a policy π such that $V_{\pi^n} \uparrow V_{\pi}$.*

Proof. Suppose that $\{V_{\pi^n}\}_{n=1}^{\infty}$ be generated by soft-policy iteration. By Lemma 12, the sequence of value functions $\{V_{\pi^n}\}_{n=1}^{\infty}$ is non-decreasing and bounded so $\{Q_{\pi^n}\}_{n=1}^{\infty}$ is non-decreasing and bounded as well. Therefore, the monotone convergence theorem give us

$$\lim_{n\to\infty} Q_{\pi^n}(s,a) = r(s,a) + \int_{\mathcal{S}} V(s')\rho_{(s,a)}(ds') = Q(s,a),$$

where V is defined in Lemma 12. Now, by Schäl (1974, Lemma 4)[1] there exists a policy π such that for every state s in $\mathcal{S}$, π_s is an accumulation point of $\{\pi_s^n\}_{n=1}^{\infty}$. Therefore, for a fixed state s we can find a subsequence $\{\pi_s^{n_k}\}_{k=1}^{\infty}$ converging weakly to π_s. We claim that $V_{\pi} = V$.

We begin by showing that $V \leq V_{\pi}$. Since $V_{\pi^n} \uparrow V$ for $n \to \infty$ we have

$$\begin{aligned} V(s) &= \lim_{n\to\infty} V_{\pi^n}(s) = \lim_{k\to\infty} V_{\pi^{n_k}}(s) \\ &\leq \limsup_{k\to\infty} \int_{\mathcal{A}} Q_{\pi^{n_k}}(s,a)\pi_s^{n_k}(da) - \tau \liminf_{k\to\infty} \mathrm{KL}[\pi_s^{n_k}||\bar{\pi}_s]. \end{aligned}$$

From Dupuis & Ellis (2011, Lemma 1.4.3) the KL-divergence is lower semicontinuous so

$$\mathrm{KL}[\mu||\bar{\pi}_s] \leq \liminf_{n\to\infty} \mathrm{KL}[\mu_n||\bar{\pi}_s]. \tag{5}$$

On the other hand, Feinberg et al. (2014, Theorem 1.1) shows that

$$\limsup_{n\to\infty} \int_{\mathcal{A}} Q_{\pi^{n_k}}(s,a)\pi_s^{n_k}(da) \leq \int_{\mathcal{A}} Q(s,a)\pi_s(da) \tag{6}$$

Then combining (5) and (6) we have

$$V(s) \leq \int Q_{\pi}(s,a)\pi_s(da) - \tau \mathrm{KL}[\pi_s||\bar{\pi}_s] = [\mathcal{T}_{\pi}V](s).$$

Therefore π is proper by Lemma 10 and $V_{\pi} \geq V$. Now we show the reverse inequality. From the proof of Lemma 12 we have

$$\mathcal{T}_{\pi}V_{\pi^n} \leq \mathcal{T}V_{\pi^n} \leq V \text{ for all } n \text{ in } \mathbb{N}.$$

By taking the limit as $n \to \infty$ and making use of the monotone convergence theorem and that $V_{\pi_n} \uparrow V$ we get $\mathcal{T}_{\pi}V \leq V$. Then from Lemma 7 it follows that

$$V_{\pi} = \lim_{n\to\infty} \mathcal{T}_{\pi}^n V \leq V,$$

resulting in $V_{\pi} = V$. □

Lemma 14. *The operator $\mathcal{T}$ has the cost shifting property i.e. if $V : \mathcal{S} \to \mathbb{R}$ is a bounded measurable function and $\delta \geq 0$, then $\mathcal{T}V + \delta\mathbb{1} \geq \mathcal{T}(V + \delta\mathbb{1})$.*

Proof. Follows from a simple calculation using the definition of $\mathcal{T}$. □

[1] with $\mathcal{A}$ replaced by $\mathcal{P}(\mathcal{A})$ and noting that $\mathcal{P}(\mathcal{A})$ is compact.

Lemma 15. *Let Assumptions 1 and 2 hold and suppose that the optimal value function is bounded above. Let π be a policy such that soft-policy iteration, starting at any proper policy, converges to V_π. Then soft-value iteration, starting from any bounded measurable function, also converges to V_π.*

Proof. Let V be a bounded measurable function. We want to show that $\mathcal{T}^n V \to V_\pi$ as $n \to \infty$. To show this we will sandwich $\mathcal{T}^n V$ between to sequences that converge to V_π. First we construct the lower sequence in the sandwich. Let $\delta > 0$. Then there is a unique bounded measurable function V_δ such that

$$V_\delta + \delta \mathbb{1} = \mathcal{T}_\pi V_\delta.$$

To see this, note that since π is proper, it must also be proper for a shifted problem where the reward is decreased by δ so the equation $V = \mathcal{T}_\pi V - \delta \mathbb{1}$ must have a unique (bounded and measurable) solution. Additionally, it is clear from this argument that $V_\delta \leq V_\pi$. In view of this and the monotonicity of $\mathcal{T}$ we have

$$V_\pi = \mathcal{T} V_\pi \geq \mathcal{T} V_\delta \geq \mathcal{T}_\pi V_\delta = V_\delta + \delta \mathbb{1} \geq V_\delta.$$

Using the monotonicity of $\mathcal{T}$ again it follows that

$$V_\pi = \mathcal{T}^n V_\pi \geq \mathcal{T}^n V_\delta \geq \mathcal{T}^{n-1} V_\delta \geq V_\delta \text{ for all } n \text{ in } \mathbb{N}.$$

Therefore, $\{\mathcal{T}^n V_\delta\}_{n=1}^\infty$ is a non-decreasing sequence of functions bounded above by V_π. Hence $\mathcal{T}^n V_\delta \uparrow \hat{V}_\delta$ for some bounded measurable function $\hat{V}_\delta$. It follows, by Lemma 6, that

$$\hat{V}_\delta = \lim_{n \to \infty} \mathcal{T}^n \hat{V}_\delta = \mathcal{T} \hat{V}_\delta,$$

so $\hat{V}_\delta$ satisfies the optimality equation. But V_π is the unique solution to the optimality equation so $\hat{V}_\delta = V_\pi$ and $\mathcal{T}^n V_\delta \uparrow V_\pi$. Next, we construct the upper sequence in the sandwich. Together with the cost shifting property (Lemma 14), the monotonicity of $\mathcal{T}$ gives

$$V_\pi + \delta \mathbb{1} = \mathcal{T} V_\pi + \delta \mathbb{1} \geq \mathcal{T}(V_\pi + \delta \mathbb{1}) \geq \mathcal{T} V_\pi = V_\pi.$$

Hence $\{\mathcal{T}^n(V_\pi + \delta \mathbb{1})\}_{n=1}^\infty$ is a non-increasing sequence of functions bounded below by V_π. Then the same argument used above shows that $\mathcal{T}^n(V_\pi + \delta \mathbb{1}) \downarrow V_\pi$. Now, since $V_\delta \leq V_\pi$ and π is proper we have

$$V_\delta = \mathcal{T}_\pi V_\delta - \delta \mathbb{1} \leq \mathcal{T}_\pi V_\pi - \delta \mathbb{1} = V_\pi - \delta \mathbb{1} < V_\pi + \delta \mathbb{1}.$$

Thus, we can find a $\delta > 0$ such that V lies between V_δ and $V_\pi + \delta \mathbb{1}$. As a result, the monotonicity of $\mathcal{T}$ implies that

$$\mathcal{T}^n V_\delta \leq \mathcal{T}^n V \leq \mathcal{T}^n (V_\pi + \delta \mathbb{1}).$$

Since $\mathcal{T}^n V_\delta \to V_\pi$ and $\mathcal{T}^n(V_\pi + \delta \mathbb{1}) \to V_\pi$ the proof is complete. □

Theorem 1. *Let Assumptions 1 and 2 hold and suppose that the optimal value function is bounded above. Then (i) there exists an optimal stationary policy, (ii) the optimal value function is the unique solution to the Bellman optimality equation (iii) soft value iteration converges to the optimal value function starting from any bounded function; and (iv) soft policy iteration converges to the optimal policy starting from any proper policy.*

Proof. From Lemmas 12, 13 and 15 we have that there exists a policy π such that (a) V_π is the unique solution to the optimality equation; (b) soft-value iteration converges to V_π starting from any bounded function; and (c) soft-policy iteration converges to V_π starting from any proper policy. Finally, we show that π is an optimal policy. Let π' be an arbitrary policy. Then we have $\mathcal{T}_{\pi'}^n 0 \leq \mathcal{T}^n 0$. Then taking the limit superior as $n \to \infty$ gives $V_{\pi'} \leq V_\pi$. Since π' was arbitrary $V_\pi(s) = \sup_{\pi'} V_{\pi'}(s)$ for all states s in $\mathcal{S}$. □

3. DQN Architecture and Hyperparameters

In our experiments, we used a DQN with the following architecture:

1. Three convolutional layers:
   (a) Layer 1 has 3 input channels, 32 output channels, a kernel size of 8 and a stride of 4.
   (b) Layer 2 has 32 input channels, 64 output channels, a kernel size of 4 and a stride of 2.
   (c) Layer 3 has 64 input channels, 64 output channels, a kernel size of 3 and a stride of 1.
2. Two fully-connected linear layers:
   (a) Layer 1 has input size 3136 and output size 512 and uses a ReLU activation function.
   (b) Layer 2 has input size 512 and output size 4 with no activation function.

We used the ADAM optimiser with batch size 32 and a learning rate of 10^{-4}. We trained every 4 timesteps and update the target Q-network every 1000 steps. Finally we used ε-greedy exploration, annealing ε to 0.01 over 100000 timesteps.

References

Dupuis, P. and Ellis, R. *A weak convergence approach to the theory of large deviations*. 2011.

Feinberg, E. A., Kasyanov, P. O., and Zadoianchuk, N. V. Fatou's lemma for weakly converging probabilities. *Theory of Probability & Its Applications*, 58:683–689, 2014.

Hernández-Lerma, O. and Muñoz de Ozak, M. Discrete-time Markov control processes with discounted unbounded costs: optimality criteria. *Kybernetika*, 28(3):191–212, 1992.

Hinderer, K. Foundations of non-stationary dynamic programming with discrete time parameter. In *Lecture Notes in Operations Research and Mathematical Systems*, volume 33. 1970.

Rudin, W. *Real and Complex Analysis*. McGraw-Hill, Inc., New York, NY, USA, 3 edition, 1987. ISBN 0070542341.

Schäl, M. A selection theorem for optimization problems. *Archiv der Mathematik*, 25:219–224, 1974.