Tight Value Interpretable Dynamic Treatment Regimes

Anonymous Author(s)
Affiliation
Address
`email`

1 VIDTR recall

Given modeling information ($P_t, r_t \forall t \in [T]$) for a Markov Decision Process, how do we obtain policies that look like decision trees?

We recall the Value Interpretable Dynamic Treatment Regimes algorithm to perform this.

At time-step T, the following is true:

$$\pi^*(s) = \begin{cases} a_1 & \text{if } s \in R_1, \\ \arg\max_a [r_T(s, a)] & \text{if } s \notin R_1 \end{cases}$$

The total loss for time-step T and length-step 1 is given by:

$$T\epsilon_T^1 := \min_{R_1, a_1} [\int_{R_1} (-r_T(s, a_1) + \max_a [r_T(s, a)])ds - \eta_T V(R_1) + \rho_T c(R_1)] \tag{1}$$

The optimal regions R_t^l are chosen as the argmin of 1.

For the kth length-step we have the piecewise continuous policy given by:

$$\pi_k^*(s) = \begin{cases} a_1 & \text{if } s \in R_1, \\ a_2 & \text{if } s \in R_2, \\ ..., & \\ a_k & \text{if } s \in R_k - \cup_{i=1}^{k-1} R_i, \\ \arg\max_a [r_T(s, a)] & \text{if } s \in S - \cup_{i=1}^{k} R_i, \end{cases}$$

The total loss for the timestep T and lengthstep k here is given by:

$$T\epsilon_T^k := \min_{R_k, a_k} [\int_{R_k - \cup_{i=1}^{k-1}} (-r_T(s, a_k) + \max_a [r_T(s, a)])ds - \eta_T V(R_k) + \rho_T c(R_k)] \tag{2}$$

At the final lengthstep l_T^*, we have:

$$\min_{a_{l_T^*}} [\int_{S - \cup_{i=1}^{l_T^* - 1} R_i} [-r_T(s, a_{l_T^*}) + \max_a (r_T(s, a))]ds - \eta_T V(S - \cup_{i=1}^{l_T^*} R_i)] \tag{3}$$

For a different time step $t < T$, we do the following replacements:

$$r_T(s, a_i) \leftrightarrow r_t(s, a_i) + \gamma P_t^{a_i} V_{t+1}(s), \ \max_a [r_T(s, a)] \leftrightarrow \max_a [r_t(s, a) + \gamma P_t^a V_{t+1}(s)] \tag{4}$$

At the end of the algorithm what we have are region-wise constant policies, given by $[(R_t^l, a_t^l)]$. Here $t \in [T]$ and $l \in [l_t^*]$. The computationally demanding steps in the method is to find the argmin for each and every length and time-step. In the next section, we assume that the rewards and transitions are close to each other in time t and thus derive approximations to the existing method.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

2 Approximation to the VIDTR based on tightness of the transitions P_t and the rewards r_t

Assume that $||P_t - P_{t+1}|| < \epsilon_P$ and $||r_t - r_{t+1}|| < \epsilon_R$, where we measure the norm by looking at the differences in all the entries. For this section, we assume that $l_k^* = L \forall k$. Here we give the computation of the argmin. We recall the computation of optimal regions and actions.

$$(R_t^l, a_t^l) := \arg\min_{R_t^l, a_t^l} [\int_{R_t^l - \cup_{i<l} R_t^i} (\max_a [r_t(s,a) + \gamma P_t^a V_{t+1}(s)] -$$
$$[r_t(s, a_t^l) + \gamma P_t^{a_t^l} V_{t+1}(s)])ds - \eta_t V(R_t^l - \cup_{i<l} R_t^i) + \rho_t c(R_t^l)]$$

We define the following operator:

$$\Psi_t(R, a|r, P) := \int_{R_t^l - \cup_{i<l} R_t^i} (\max_a [r_t(s,a) + \gamma P_t^a V_{t+1}(s)] - [r_t(s, a_t^l) + \gamma P_t^{a_t^l} V_{t+1}(s)])ds -$$
$$\eta_t V(R_t^l - \cup_{i<l} R_t^i) + \rho_t c(R_t^l)$$

We then have the optimal actions and regions given by:

$$(R_t^{l*}, a_t^{l*}) := \arg\min_{R_t^l, a_t^l} \Psi(R_t^l, a_t^l | r_t, P_t) \tag{5}$$

In the next subsection, we list various theorems that follow as a consequence of the tightness assumption of the prob. kernels and rewards.

2.1 Theorems and lemmas

In this subsection we list the main theorems used in the text: The assumption we have is tightness of the dynamics and the reward kernels i.e. $||P_t - P_{t+1}|| < \epsilon_P$ and $||r_t - r_{t+1}|| < \epsilon_R$.

Theorem 2.1. *We have the following theorem that relates how the value functions for the different times relate to each other:*

$|V_{T-i} - V_{T-i-1}| \leq C \frac{1-\gamma^{i+1}}{1-\gamma}$

Here we can see that $C := \epsilon_r + 2\gamma\epsilon_P R$

From this we can see that:

$$||V_{T-i}|| < C \frac{1-\gamma^{i+1}}{1-\gamma} + R \tag{6}$$

We also have the following theorem:

Theorem 2.2.

$$||\Psi_{t+1} - \Psi_t|| \leq |R|[3\epsilon_R + \gamma C \frac{1-\gamma^{T-t-1}}{1-\gamma}] \tag{7}$$

If we have a uniform continuity of Ψ_t for the different time steps $t \in [T]$. We can see that the minimums of $\Psi_t(R, a|r, P)$ are also close by. This step can also be useful in showing that the argmins of Ψ are close, i.e., R_{t+1}^{l*} lies in a region around R_t^{l*}.

2.2 Proofs

Theorem (2.2)
We have tightness in the transition kernels and in the rewards. This means that $d_R(r_t, r_{t+1}) < \epsilon_R$ and we have $d_P(P_t^a, P_{t+1}^a) < \epsilon_P$. How do we show that Ψ_t and Ψ_{t+1} are close? Recall that Ψ_t and Ψ_{t+1} are given by:

$$\Psi_t(R,a|r,P) := \int_{R_t^l - \cup_{i<l} R_t^i} (\max_a (r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - [r_t(s,a_t^l) + \gamma P_t^{a_t^l} V_{t+1}(s)])ds$$
$$-\eta_t V(R_t^l - \cup_{i<l} R_t^i) + \rho_t c(R_t^l)$$

$$\Psi_{t+1}(R,a|r,P) := \int_{R_{t+1}^l - \cup_{i<l} R_{t+1}^i} (\max_a (r_{t+1}(s,a) + \gamma P_{t+1}^a V_{t+2}(s))$$
$$-[r_{t+1}(s,a_{t+1}^l) + \gamma P_{t+1}^{a_{t+1}^l} V_{t+2}(s)])ds - \eta_{t+1} V(R_{t+1}^l - \cup_{i<l} R_{t+1}^i) + \rho_{t+1} c(R_{t+1}^l)$$

Let us start with length $l = 1$:

$$\Psi_{t+1}(R,a|r,P) - \Psi_t(R,a|r,P) := \int_R (\max_a [r_{t+1}(s,a) + \gamma P_{t+1}^a V_{t+2}(s)] -$$
$$\max_a [r_t(s,a) + \gamma P_t^a V_{t+1}(s)])ds + \int_R -[(r_{t+1} - r_t)(s,a) + \gamma(P_{t+1}^a V_{t+2} - P_t^a V_{t+1})(s)]ds$$

Here we get:

$$||\Psi_{t+1} - \Psi_t|| \leq 2\epsilon_R |R| + 2\gamma\epsilon_R + |\eta_t - \eta_{t+1}| V(R) + |\rho_t - \rho_{t+1}| c(R)$$
$$+ \int_R \max_a [(r_{t+1} - r_t) + \gamma(P_{t+1}^a V_{t+2} - P_t^a V_{t+1}(s))]ds$$

Here we split the above equation:

$$P_{t+1}^a V_{t+2} - P_t^a V_{t+1} = P_{t+1}^a V_{t+2} - P_{t+1}^a V_{t+1} + P_{t+1}^a V_{t+1} - P_t^a V_{t+1} \tag{8}$$

The above equation then yields the above inequality:

$$||\Psi_{t+1} - \Psi_t|| \leq |R|[3\epsilon_R + C\gamma \frac{1-\gamma^{T-t-1}}{1-\gamma}] \tag{9}$$

We have thus shown that:

$$|\Psi_{t+1}(R,a|r,P) - \Psi_t(R,a|r,P)| \leq |R|[3\epsilon_R + C\gamma \frac{1-\gamma^{T-t-1}}{1-\gamma}] \tag{10}$$

We need to now establish conditions when the argmins of two functions are close if the functions are close.

Theorem (2.1)

$$|V_{t+1}(s) - V_t(s)| = |\max_a f_{t+1}(s,a) - \max_a f_t(s,a)|$$
$$\leq \max_a |f_{t+1}(s,a) - f_t(s,a)|$$

From here we get:

$$\max_a [(r_{t+1}(s,a) - r_t(s,a)) + \gamma(P_{t+1}^a V_{t+2}(s) - P_t^a V_{t+1}(s))] \tag{11}$$

We then see:

$$\leq \epsilon_R + \gamma|V_{t+2} - V_{t+1}| + \gamma|P_{t+1}^a - P_t^a| V_{t+1} \tag{12}$$

From here we see:

$$\leq \epsilon_R + 2\gamma\epsilon_P V_{t+1} + \gamma|V_{t+2} - V_{t+1}| \tag{13}$$

At timestep T, we see that:

$$|V_T - V_{T-1}| \leq \epsilon_R + 2\gamma\epsilon_P |V_T| \tag{14}$$

For a general i, we have:

$$|V_{T-i} - V_{T-i-1}| \leq C\frac{1-\gamma^{i+1}}{1-\gamma} \tag{15}$$

We state the following theorem.

Lemma 2.3. *If $\alpha \in \mathcal{A}, \beta \in \mathcal{B}$ and we have that*

$$|f(\alpha|\beta) - g(\alpha|\beta)| \leq C \tag{16}$$

Then we can see that the following holds $\forall \alpha \in \mathcal{A}, \beta \in \mathcal{B}$:

$$|\min_{\beta \in \mathcal{B}} f(\alpha|\beta) - \min_{\beta \in \mathcal{B}} g(\alpha|\beta)| \leq C \tag{17}$$

Proof. Assume that $m_f := \min_\beta f(\alpha|\beta)$ and assume that $m_g := \min_\beta g(\alpha|\beta)$.
Let us fix an arbitrary $\alpha_0 \in \mathcal{A}$, then $\exists\, \beta' \in \mathcal{B}$ such that:

$$m_f \leq f(\alpha_0|\beta') < m_f + \epsilon \tag{18}$$

$$m_f - g(\alpha_0|\beta') \leq f(\alpha_0|\beta') - g(\alpha_0|\beta') < m_f + \epsilon - g(\alpha_0|\beta') \tag{19}$$

We then have:

$$m_f - g(\alpha_0|\beta') \leq C \tag{20}$$

We then have:

$$g(\alpha_0|\beta') \geq m_f - C \tag{21}$$

We also have:

$$m_g \geq m_f - C \tag{22}$$

This implies that:

$$m_f - m_g \leq C \tag{23}$$

Exchanging the role of f and g, we see that:

$$m_g - m_f \leq C \tag{24}$$

This gives us that:

$$|m_f - m_g| \leq C \tag{25}$$

□

Under what conditions can we show that the arg mins of the functions are close to one another? Further, is it true that the arg mins of the functions are unique?

Lemma 2.4. *Assume $f : \mathbb{R}^d \to \mathbb{R}$ to be convex, Lipschitz with constant C, i.e., $||f(x) - f(y)|| \leq C||x - y||$, and differentiable. In addition, assume $|\partial_{x_i} f| \leq D$ for all $x \in \mathbb{R}^d$. We then have that*

$$dD(x - y) \leq ||f(x) - f(y)|| \tag{26}$$

Proof. We find that the following holds:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{27}$$

When we move a part of the RHS to the left-hand side, we see that:

$$f(tx + (1-t)y) - f(y) \leq t(f(x) - f(y)) \tag{28}$$

Dividing by t on the LHS and having $t \to 0$, we see:

$$\frac{f(tx + (1-t)y) - f(y)}{t} \leq (f(x) - f(y)) \tag{29}$$

If we have $\lim_{t \to 0}$ on the LHS, what we get is:

$$(x - y)|\partial_{x_i} f| \leq (f(x) - f(y)) \tag{30}$$

From here we get the implied inequalities by looking at the bounds over $\partial_i f \leq D$ and summing the LHS inequality d times. □

Lemma 2.5. *Assume f, g to be convex. Lipschitz with constant C, further assume that the derivatives are bounded by D. Also, assume that f and g are uniformly close by ϵ. Assume x_f is the minimizer of f, and x_g is the minimizer of g. In this case, we can show that x_f and x_g are close:*

$$||x_f - x_g|| \leq \frac{2\epsilon}{dD} \tag{31}$$

Proof. We have the following to be true:

$$||x_f - x_g|| \leq \frac{1}{dD}||f(x_f) - f(x_g)|| \tag{32}$$

Adding and subtracting $g(x_g)$, we get:

$$||f(x_f) - f(x_g)|| \leq ||f(x_f) - g(x_g) + g(x_g) - f(x_g)|| \tag{33}$$

Further splitting of the above equation yields:

$$||f(x_f) - f(x_g)|| \leq ||f(x_f) - g(x_g)|| + ||g(x_g) - f(x_g)|| \tag{34}$$

Since g, f are uniformly close, which also implies that the minimums are uniformly close, what we see is that:

$$||x_f - x_g|| \leq \frac{2\epsilon}{dD} \tag{35}$$

□

To show the convexity of Ψ_t, we need to show that the following holds:

$$\begin{aligned}
&\int_{\lambda a_1+(1-\lambda)a_2}^{\lambda b_1+(1-\lambda)b_2}[\max_a(r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - (r_t(s,a) + \gamma P_t^{a'} V_{t+1}(s))]ds \\
&\leq \lambda[\int_{a_1}^{b_1} \max_a(r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - (r_t(s,a) + \gamma P_t^{a'} V_{t+1}(s))ds] + \\
&(1-\lambda)[\int_{a_2}^{b_2} \max_a(r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - (r_t(s,a) + \gamma P_t^{a'} V_{t+1}(s))ds]
\end{aligned}$$

What about the boundedness of the derivatives? Is it true that,

$$\frac{\partial}{\partial_b}\int_a^b [\max_a(r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - (r_t(s,a) + \gamma P_t^a V_{t+1}(s))]ds \leq C \text{ for some } C > 0 \tag{36}$$

We can see that this holds if we assume that $|r_t(s,a)| \leq R$, and $|P_t^a| \leq 1$, Total Reward $\leq 1 + R + R^2 + ... + R^{T-1} \leq \frac{R^T - 1}{R-1}$ and thus the value function $V_t(s) \leq \frac{R^t-1}{R-1}$. This implies that we can bound the derivatives with:

$$\begin{aligned}
&|\max_a(r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - (r_t(s,a) + \gamma P_t^a V_{t+1}(s))| \\
&\qquad \leq 2(R + \gamma\frac{R^{t+1}-1}{R-1})
\end{aligned}$$

For convexity it would be enough to show that the below function has positive Hessian matrix:

$$\Psi_t(x, a|r, P) := \int_{s_0}^x [\max_a(r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - (r_t(s,a) + \gamma P_t^a V_{t+1}(s)) - \eta ds] \tag{37}$$

The double derivative of the above function is:

$$\max_a[\partial_s r_t(s,a) + \gamma P_t^a V'_{t+1}(s)] - (\partial_s r_t(s,a) + \gamma P_t^a V'_{t+1}(s)) \tag{38}$$

How and when can we assume that this is positive?

How do we answer the following question?

Theorem 2.6. *Given a function f, how do we find a convex function g such that we have to solve $\min_g \int_D (f(x) - g(x))dx$ with $\nabla^2 g \succ 0$. Further, can we characterize the minimal error $\int_D (f(x) - g(x))dx$? D here denotes the domain of the function f.*

Here we list a sequence of questions and answers about the convexity condition we wish to answer for functions:

1. Is local convexity almost everywhere enough to assume inverse Lipschitzness? No, take the example of a piecewise linear function that has a peak.
2. What about local convexity? Yes, this is enough to show that function is globally convex provided twice differentiability. We find that if f' is non-decreasing, then f is convex. However, we need to show this at all points in the domain in order to get the proof to go through.

Q) Can we compute the errors in approximation; atleast for the 1 dimensional case? The problem we wish to solve here is:

$$\min_g \int_D |f(x) - g(x)| dx \tag{39}$$

We have the following condition that we subject our problem to:

$$\nabla^2 g \succ 0 \tag{40}$$

In the 1-D case, this problem can be restated as:

$$\min_g \sum_{i=0}^{N} ||f(x_i) - g(x_i)||_{2/1} \tag{41}$$

This is subject to the following condition:

$$g(x_{i+1}) + g(x_{i-1}) - 2g(x_i) \geq 0 \tag{42}$$

We solve a general version of the problem given by the following. We solve the matrix positive hyperplane problem for a given vector f and matrix A. The problem is defined as follows: Given f values on the $\mathbb{R}^1$ a vector of length $(c, 1)$ and a matrix A of dimensions (r, c) of full-row rank. Solve the following problem:

$$\min_g ||f - g|| \tag{43}$$

Subject to:

$$Ag \succ 0 \tag{44}$$

Compute the SVD of $A := U\Sigma V^T$, we then have the following condition $U\Sigma V^T g \succ 0$. Define $y := V^T g$. Left multiplying by V^T in the optimization problem 43. We get the following modified problem

$$\min_y ||V^T f - y|| \tag{45}$$

Assuming that singular values are $\lambda_1, \lambda_2, ..., \lambda_r$. The condition which this is subject to is given by:

$$U[\lambda_1 y_1, \lambda_2 y_2, ..., \lambda_r y_r]^T \succ 0 \tag{46}$$

Since $\lambda_1, \lambda_2, ..., \lambda_r > 0$, we get that this condition then gives us $(y_1, y_2, ..., y_r)^T \in U^{-1}\mathbb{H}^r_+$. To the optimization problem 45 we left multiply by $[[U, O_{r,c-r}], [O_{c-r,r}, I_{c-r,c-r}]]$ where O is a zero matrix with the dimensions in the subscript. We can do this and the norm does not change since the above matrix is orthogonal since each block is orthogonal. We then get the following in the norm:

$$||[[U, O], [O, I]]^T V^T f - [U\Pi_r y, y_{r+1}, ..., y_c]^T|| \tag{47}$$

Here and elsewhere we have that $\Pi_i y = [y_1, y_2, ..., y_i]^T$, where $y = [y_1, y_2, ..., y_n]^T$
This is subject to

$$\Pi_r y \in U^{-1}\mathbb{H}^r_+ \tag{48}$$

After left multiplying by U, we see that this splits as:

$$||[U\Pi_r V^T f, \Pi_{c-r} V^T f]^T - [U\Pi_r y, y_{r+1}, y_{r+2}, ..., y_c]^T|| \tag{49}$$

The above 49 then splits as:

$$||U\Pi_r V^T f - U\Pi_r y|| + ||\Pi_{c-r} V^T f - \Pi_{c-r} y|| \tag{50}$$

This is subject to $\Pi_r y \in U^{-1}\mathbb{H}_+^r$
Since there are no conditions on $\Pi_{c-r}y$, we can have $\Pi_{c-r}y = \Pi_{c-r}V^T f$, this then implies that the second part of 49 can be taken to be 0 which leaves us with:

$$||U\Pi_r V^t f - U\Pi_r y|| \tag{51}$$

Here we must have that $U\Pi_r y \in \mathbb{H}_+^r$ and this can be any r dimensional positive vector. This implies we can choose $(U\Pi_r y)_i = (U\Pi_r V^t f)_i$ if $(U\Pi_r V^t f)_i > 0$, else we must have $(U\Pi_r y)_i = 0$.

$$\begin{aligned}(U\Pi_r y)_i &= (U\Pi_r V^t f)_i \text{ if } (U\Pi_r V^t f)_i \geq 0 \\ &= 0 \text{ if } (U\Pi_r V^t f)_i < 0\end{aligned} \tag{52}$$

Let P_n be a positive vector map from $\mathbb{R}^n \to \mathbb{R}^n$ such that if $v_i \geq 0$ then $(P_n v)_i = v_i$ and if $v_i < 0$ then $(P_n v)_i = 0$.

$$\begin{aligned}(P_n v)_i &= v_i \text{ if } v_i \geq 0 \\ &= 0 \text{ if } v_i < 0\end{aligned} \tag{53}$$

Then we get the following to be true for $\Pi_r y = U^{-1}P_r(U\Pi_r V^t f)$ and we must have the norm to be given by $||(I - P_r)U\Pi_r V^t f||$. We formally state the theorem as follows:

Theorem 2.7. *Given a vector f of dimensions $(c, 1)$ and a matrix A of dimension (r, c) of full row-rank the solution to the following constrained optimization problem:*

$$\min_g ||f - g|| \tag{54}$$

Subject to:

$$Ag \succ 0 \tag{55}$$

If we assume $y := V^t g$, then we see that the solution is given by

$$\Pi_r y = U^t P_r(U\Pi_r V^t f) \tag{56}$$

and we also have:

$$\Pi_{c-r}y = \Pi_{c-r}(V^t f) \tag{57}$$

where Π_r is the projection map from $\mathbb{R}^c \to \mathbb{R}^r$ and P_r is the positivity map mapping $(v_i)_{i=1}^r \to (\mathcal{I}(v_i > 0)v_i)_{i=1}^r$. The minimizing norm is given by $||(I - P_r)U\Pi_r V^t f||$.

In this section we state the two dimensional version of the convex approximation problem and find ways to solve the problem. The problem is given as following:

$$\min_{g} ||f - g|| \tag{58}$$

Subject to the $2 - D$ condition which is given by:

$$\nabla^2 g \succ 0 \tag{59}$$

The two dimensional derivative is given by $\begin{pmatrix} \partial^2 g/\partial x_1^2 & \partial^2 g/\partial x_1 \partial x_2 \\ \partial^2 g/\partial x_1 \partial x_2 & \partial^2 g/\partial x_2^2 \end{pmatrix}$ The constraint implies that the principal minors of $\nabla^2 g \succ 0$. This implies that the $(1, 1)$th entry is positive and the determinant of the Hessian is positive. We now split and write the numerical approximation of each term in the derivative:

$$\frac{g(x_1 + h, x_2) + g(x_1 - h, x_2) - 2g(x_1, x_2)}{h^2} \geq 0 \tag{60}$$

The determinant is then given by:

$$[\frac{g(x_1 + h, x_2) + g(x_1 - h, x_2) - 2g(x_1, x_2)}{h^2}][\frac{g(x_1, x_2 + h) + g(x_1, x_2 - h) - 2g(x_1, x_2)}{h^2}] \geq$$
$$[\frac{g(x_1 + h, x_2 + h) + g(x_1 - h, x_2 - h) - g(x_1 + h, x_2 - h) - g(x_1 + h, x_2 - h)}{4h^2}]^2$$

Borkar's paper for approximating the convex envelope for a function. In the paper, they present a Q-Learning method to approximate the convex envelope for a function.

For a function $f : \mathbb{R}^d \to \mathbb{R}$, we define its convex

Q) Can you show that A has strictly positive singular values for all values N?
Ans: Yes, singular values for a matrix are positive by definition and A is a surjection.

Q) Can you plot the errors in the $1-D$ convex optimization problem for the Ψ function in CVXPY?
Ans: Yes, the function and it's approximation seem quite close.

Q) Can you verify the maths for the $1-D$ approximation problem?
Ans: Yes, done.

Q) Can you make the connection to our problem and derive the corresponding solution?
Ans: Yes, in our problem $A = [[1, -2, 1, 0..., 0], [0, 1, -2, -1, ..., 0], ..., [0, 0, ..., 1, -2, 1]]$, where the dimension of A is $(N-1, N+1)$ and the full row rank.

Q) Can you code up your solution and compare it with the one in the CVXPY library?
Ans: Yes, the errors are off.

Q) Can you formulate the convex optimization problem for dimensions 2 and above? Can you solve these analytically or computationally?
Ans: Yes, I can formulate the problem; but it has many constraints and is multidimensional optimization.

Q) Read and understand the duality-based approach to convexity.
Ans:

Q) Can we extend the theorem to d dimensional space?
Ans:

Q) Read the approach to perform the quasi-convex approximation.
Ans:

Q) Read the paper to find an $O(n)$ algorithm to find the approximation.
Ans:

Q) Read the paper sent by Agni on quasi-convex approximations/ convex envelopes.
Ans: The paper gives a method for constructing a quasi-convex approximation and characterizes the errors in the approximations under certain constraints on the function to be approximated.
Subquestions:
a) Can we implement this on a computer?
b) What are the conditions and can we assume this holds for the Ψ function?

Q) Does this solve the convex optimization problem, at least for the $1-D$ space?
Ans:

Q) Can you code up the duality-based solution for $1D$?
Ans:

Q) Can you solve the convex optimization problem, for dimension d space based on duality?
Ans:

Q) Can you code the solution for a d dimensional space?
Ans:

Q) Can you edit and rewrite the script with theorems, proofs, and other bits?
Ans:

Q) Can you finish the VIDTR proof? The last step is left.
Ans: