Lipschitz Inner Functions in Kolmogorov's Superposition Theorem

Jonas Actor

Rice University

13 September 2017

Table of Contents

Table of Contents

Hilbert's 13th Problem

Problem Statement (1900)

Can the solution x to the 7th degree polynomial equation

$$x^7 + ax^3 + bx^2 + cx + 1 = 0$$

be represented by a finite number of compositions of bivariate continuous/analytic functions using the three variables a, b, c?

Hilbert's 13th Problem

Larger focus to Hilbert's Problem:

- Which functions can be represented using a finite number of compositions of simpler functions?
  - Continuous functions $\to$ Continuous functions?
  - Analytic functions $\to$ Continuous functions?
  - Analytic functions $\to$ Analytic functions?
- How 'simple' or 'complex' is a function?
  - Number of variables?
  - Some other measure

An Example

$$f(x_1, x_2) = x_1 x_2$$

Can we create ϕ_1, ϕ_2, ψ_{11}, ψ_{12}, ψ_{21}, ψ_{22} univariate polynomials such that

$$f(x_1, x_2) = \sum_{i=1}^{2} \phi_i \circ \sum_{j=1}^{2} \psi_{ij}(x_j)?$$

An Example

Let

$$\phi_1(z) = \frac{1}{4}z^2 \qquad \phi_2(z) = -\frac{1}{4}z^2,$$

and

$$\begin{aligned} \psi_{11}(x_1) &= x_1 & \qquad \psi_{12}(x_2) &= x_2 \\ \psi_{21}(x_1) &= x_1 & \qquad \psi_{22}(x_2) &= -x_2 \end{aligned}$$

Then,

$$\begin{aligned} \sum_{i=1}^{2} \phi_i\left(\sum_{j=1}^{2} \psi_{ij}(x_j)\right) &= \frac{1}{4}(x_1 + x_2)^2 - \frac{1}{4}(x_1 - x_2)^2 \\ &= \frac{1}{4}(x_1^2 + 2x_1x_2 + x_2^2) - \frac{1}{4}(x_1^2 - 2x_1x_2 + x_2^2) \\ &= \frac{1}{4}(2x_1x_2 + 2x_1x_2) \\ &= x_1x_2. \end{aligned}$$

An Example

Can we do better?

- Use fewer terms!

What concessions do we make?

- Give up using polynomials, instead continuous functions

An Example

Let

$$\phi_1(z) = \exp(z) \qquad \psi_i(x_i) = \log(x_i)$$

Then,

$$\begin{aligned}\phi_1\left(\sum_{j=1}^{2} \psi_j(x_j)\right) &= \exp(\log(x_1) + \log(x_2)) \\ &= exp(\log(x_1 x_2)) \\ &= x_1 x_2.\end{aligned}$$

We have reduced the number of outer summands by 1.

An Example

- Representation not unique
- Number of terms in (inner) summand depends on dimension
- Traded ‘complexity’ of functions used
- Note, these are all very special functions

Towards Kolmogorov's Superposition Theorem

- Hilbert conjectured the answer was ‘no’, that even for continuous functions, such a representation was not always possible
- A. Kolmogorov and V.I. Arnold made progress in the 1950s
- Arnold (at age 19) proved the answer was ‘yes’ in 1957: any multivariate continuous function can be represented as a superposition of bivariate continuous functions
- Two weeks later, Kolmogorov reduced the bivariate functions from Arnold to univariate functions

Kolmogorov's Superposition Theorem (KST) (1957)

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R} \in C\left([0,1]^n\right)$ where $n \geq 2$. Then, there exist $\psi^{pq} : [0,1] \to \mathbb{R} \in C[0,1]$ and $\chi_q : \mathbb{R} \to \mathbb{R} \in C(\mathbb{R})$, where $p \in \{1, \ldots, n\}$ and $q \in \{0, \ldots, 2n\}$, such that

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi^q \left(\sum_{p=1}^{n} \psi^{pq}(x_p) \right).$$

This reformulation is due to Sprecher.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R} \in C\left([0,1]^n\right)$ where $n \geq 2$. Fix $\epsilon \leq \frac{1}{2n}$, and choose $\lambda \in \mathbb{R}$ such that $1 = \lambda^0, \lambda^1, \ldots, \lambda^{n-1}$ are integrally independent. Then, there exist $\psi : [-1,1] \to \mathbb{R} \in C[-1,1]$ and $\chi_q : \mathbb{R} \to \mathbb{R} \in C(\mathbb{R})$ for $q \in \{0, \ldots, 2n\}$, such that

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \chi^q \left(\sum_{p=1}^{n} \lambda^p \psi(x_p + q\epsilon) \right).$$

Implications

Multivariate functions suffer from the ‘curse of dimensionality’, making computation hard for higher dimensions.

Multivariate continuous functions are really

$$\begin{cases} \text{univariate continuous functions} \\ \text{addition} \\ \text{function composition} \end{cases}$$

We understand each of those three things very well...

Implications

Quest: Can we use KST to represent multivariate functions for efficient computation?

Table of Contents

Setup

- Constructions are by induction on $j \in \mathbb{N}$
- Throughout the rest of this talk, fix $\epsilon = 1/2n$

Space Partitioning

Town: closed interval

System of towns: set of disjoint closed intervals

- $\mathscr{T}_j$ a system of towns $\subseteq [-1, 1]$
- $\mathscr{T}_j^q$ a system of towns $\subseteq [-1 + q\epsilon, 1 + q\epsilon]$ where

$$\mathscr{T}_j^q = \{t + q\epsilon \ : \ t \in \mathscr{T}_j\}.$$

- Enumerate the towns in $\mathscr{T}_j^q$ by indices $1 \leq i \leq m_j$

$$\mathscr{T}_j^q = \{t_1^q,\ t_2^q, \ldots, t_{m_j}^q\}.$$

Space Partitioning

Squares: products of towns

- $S^q_{j;i_1,\ldots,i_n} = \prod_{p=1}^n t^q_{i_p}$ for towns $t^q_{i_p} \in \mathscr{T}^q_j$ for $p = 1, \ldots, n$
- $\mathscr{S}^q_j = \{S^q_{j;i_1,\ldots,i_n}, : \ 1 \leq i_1, \ldots, i_n \leq m_j\}$ set of all squares

Squares are pairwise disjoint: for any q, if $(i_1, \ldots, i_n) \neq (i'_1, \ldots, i'_n)$, then

$$S^q_{j;i_1,\ldots,i_n} \cap S^q_{j,i'_1,\ldots,i'_n} = \emptyset.$$

Fundamental Lemma

Define $\Psi^q(x_1, \ldots, x_n) = \sum_{p=1}^n \psi^{pq}(x_p)$ for each $q \in \{0, \ldots, 2n\}$, where $\psi^{pq} \in C[0,1]$

Lemma

For each j, q, suppose the families of squares $\mathscr{S}_j^q$ satisfy the following:

1. Each point $x \in [0,1]^n$ intersects squares from at least $n+1$ of the $2n+1$ families of squares
2. $\sup_{S^q_{j;i_1,\ldots,i_n} \in \mathscr{S}^q_j} \mathrm{Diam}[S^q_{j;i_1,\ldots,i_n}] \to 0$ uniformly as $j \to \infty$
3. $\Psi^q(S^q_{j;i_1,\ldots,i_n}) \cap \Psi^q(S^q_{j,i'_1,\ldots,i'_n}) = \emptyset$ when $(i_1, \ldots, i_n) \neq (i'_1, \ldots, i'_n)$.

Then, any function $f \in C([0,1]^n)$ admits a KST representation

$$f = \sum_{q=0}^{2n} \chi^q \circ \Psi^q.$$

Fundamental Lemma: Reformulation

- $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ integrally independent
- $\psi_j : [-1, 1] \to \mathbb{R}$
- $\psi := \lim_{j \to \infty} \psi_j$ uniformly
- For $q \in \{0, \ldots, 2n\}$ define $\Psi^q : [0, 1]^n \to \mathbb{R}$

$$\Psi^q(x_1, \ldots, x_n) = \sum_{p=1}^{n} \lambda_p \psi(x_p + q\epsilon).$$

Fundamental Lemma: Reformulation

Lemma: Reformulation

For each j, q, suppose the systems of towns $\mathscr{T}_j^q$ and function ψ_j satisfy the following:

1. Each point $x \in [0,1]$ intersects towns from at least $2n$ of the $2n+1$ systems of towns
2. $\sup_{t \in \mathscr{T}_j} \mathrm{Diam}(t) \to 0$ uniformly as $j \to \infty$
3. $\psi_j(t_1) \cap \psi_j(t_2) = \emptyset$ but are rational for any $t_1,\ t_2 \in \mathscr{T}_{j'}$ for $j' \leq j$

Then, any function $f \in C([0,1]^n)$ admits a KST representation

$$f = \sum_{q=0}^{2n} \chi^q \circ \Psi^q.$$

Comments

- Construction of squares (specifically, the $n+1$ of $2n+1$ property and the shrinking diameter) are important for outer function
- Hard part of inner function construction is the pairwise disjoint image condition

Smoothness of Inner Functions

- Original KST Proof (1957): Hölder continuous
  - Squares are uniformly spaced, and scale self-similarly
- Fridman (1967): Can be Lipschitz continuous, constant 1
- Vitushkin and Henkin (1954): Not differentiable at a dense set of points

Lipschitz vs. Hölder Continuity

Definition

A function $f : [0,1]^n \to \mathbb{R}$ is **Lipschitz continuous with constant C** if $\forall x, y \in [0,1]^n$,

$$\|f(x) - f(y)\| \leq C \|x - y\| .$$

Definition

A function $f : [0,1]^n \to \mathbb{R}$ is **Hölder(α) continuous with constant C** for $\alpha \in (0,1)$, if $\forall x, y \in [0,1]^n$,

$$\|f(x) - f(y)\| \leq C \|x - y\|^{\alpha} .$$

Hölder functions suffer from high storage/evaluation complexity, making them impractical for computation

Table of Contents

Terminology

- $\mathscr{T}_j$: system of towns (closed intervals) at refinement level j
- $\mathscr{T}_j^q$: system of shifted towns

$$\mathscr{T}_j^q = \left\{ t + q\epsilon \ : \ t \in \mathscr{T}_j, q \in \{-2n, \ldots, 2n\} \right\}.$$

- $\psi_j : [-1, 1] \to \mathbb{R}$
- $\psi = \lim_{j \to \infty} \psi_j$
- $\Xi_j : [0, 1] \to \mathbb{N}$

$$\Xi_j(x) = \left| \left\{ q \in \{0, \ldots, 2n\} \ : \ \exists t \in \mathscr{T}_j^q \text{ such that } x \in t \right\} \right|$$

Lemma

Lemma

It is sufficient to complete KST representation, if for each $j \in \mathbb{N}$, the system of towns $\mathscr{T}_j$ and the function $\psi_j : [-1,1] \to \mathbb{R} \in C[-1,1]$ satisfy the following:

1. $\sup_{t \in \mathscr{T}_j} \text{Diam}(t) \to 0$ uniformly as $j \to \infty$.
2. $\forall x \in [0,1]$, $\Xi_j(x) \in \{2n, 2n+1\}$
3. $\forall t \in \mathscr{T}_j$, $\psi_j(t) \in \mathbb{Q}$
4. $\forall t_1 \neq t_2 \in \mathscr{T}_j$, $\psi(t_1) \cap \psi(t_2) = \emptyset$.
5. ψ_j is piecewise linear with maximum slope of $\widehat{m}_j = 1 - 2^{-j}$.

Strategy

Start with $\psi_0 \equiv 0$ and $\mathscr{T}_0 = \{[-1,1]\}$. Then for $j \in \mathbb{N}$ do:

Select $\widehat{\mathscr{T}_j} \subseteq \mathscr{T}_j$ towns to break (length $\geq 2^{-j}$)

- Find Holes
- Solve for Plugs
- Create Gaps and Update

Find Holes

Break $t \in \widehat{\mathscr{T}_j}$ at its midpoint p by removing an open interval g.

$$t \mapsto t_- \cup g \cup t_+ \qquad\qquad p \in g$$

Danger: p might no longer be contained by enough systems of towns!
Might cause $\Xi_{j+1}(p) = 2n - 1$.

Find Holes

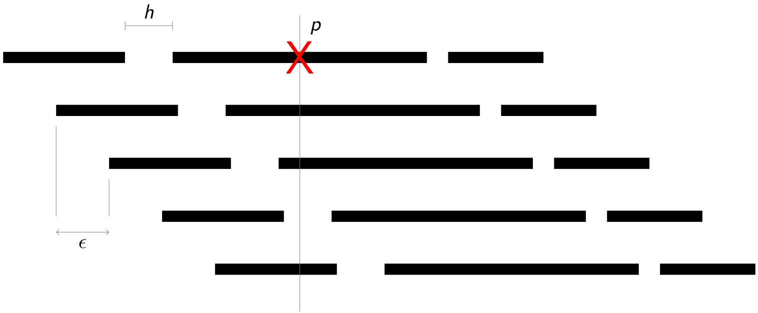

Figure: Sketch of scenario where a break point p falls into a hole, $n = 2$.

Solve for Plugs

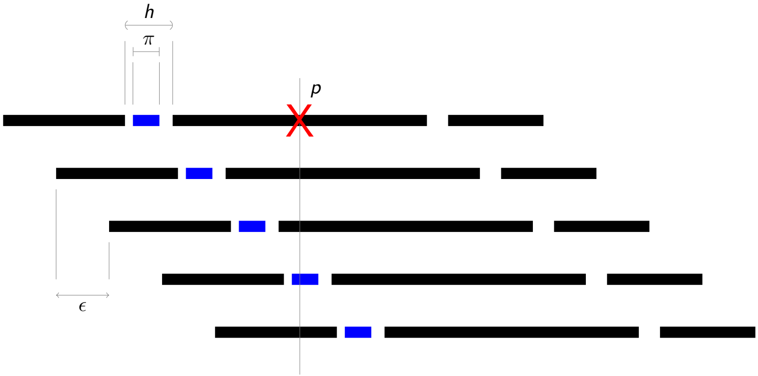

Figure: Sketch of adding a plug to our previous scenario.

Solution: Add in small 'plugs' so that when we remove a gap around p, we do not lose containment

Solve for Plugs

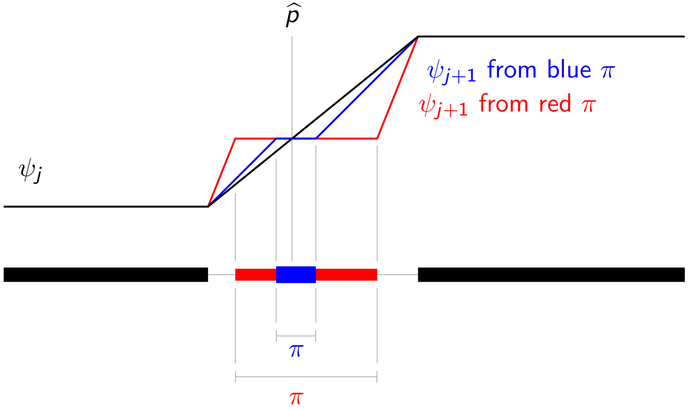

Figure: Sketch of how size of plugs changes the slope of ψ_{j+1}.

Danger: Adding in a plug $\implies$ slope of ψ_{j+1} might exceed our bound!

Solve for Plugs

How big can π be for a maximum slope of $\widehat{m} = 1 - 2^{-j-1}$?

- Might need more than one plug per break point (no more than 2)
- Might need more than one plug per hole

Solution: Solve a linear system!

Solve for Plugs

- For ν break points p_i, $1 \leq i \leq \nu$, that shift into hole $h = (b_0, a_{\nu+1})$
- Want disjoint plugs $\pi_i = [a_i, b_i]$ **$\mathbf{a_i}$, $\mathbf{b_i}$ unknown**

$$\widehat{p_i} := p_i - q_i\epsilon \in \pi_i \subset h$$

- ψ_{j+1} monotonic, piecewise linear, constant on towns/plugs
- $\psi_j(\widehat{p_i}) = \psi_{j+1}([a_i, b_i])$

We use the following notation for (known) function values:

$$\begin{aligned} f_0 &= \psi_j(b_0) \\ f_i &= \psi_j(\widehat{p_i}) \qquad 1 \leq i \leq \nu \\ f_{\nu+1} &= \psi_j(a_{\nu+1}) \end{aligned}$$

Solve for Plugs

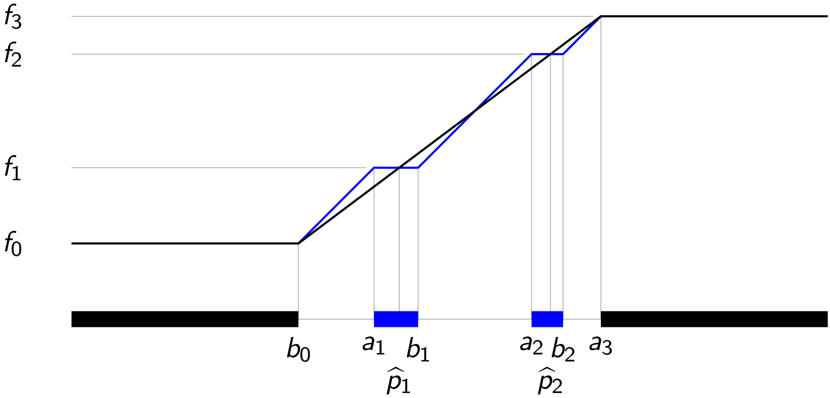

Figure: Sketch of scenario for finding two plugs, with ψ_j in black and ψ_{j+1} in blue. Note the symmetry constraint $b_1 - \widehat{p}_1 = \widehat{p}_2 - a_2$ is enforced.

Solve for Plugs

$\nu + 1$ equations:

$$\widehat{m}(a_i - b_{i-1}) = f_i - f_{i-1}, \qquad 1 \leq i \leq \nu + 1.$$

$\nu - 1$ symmetry constraints:

$$b_i - \widehat{p}_i = \widehat{p}_{i+1} - a_{i+1}, \qquad 1 \leq i \leq \nu - 1.$$

Solve for Plugs

This provides the linear system $Cx = z$

- C is block diagonal, invertible $\implies$ unique solution exists
- ψ_j monotonic increasing $\implies$ plugs are disjoint with non-empty interior

For each h, update $\mathscr{T}_j$ to include the plugs π_i.

Create Gaps

Recall our goal to 'break' $t \in \widehat{\mathscr{T}_j}$ at break point p:

$$t \mapsto t_- \cup g \cup t_+ \qquad\qquad p \in g$$

At this point, $\Xi_j(p) = 2n + 1 \implies \Xi_{j+1}(p) \geq 2n$:

$$\forall q \in \{0, \ldots, 2n\}, \ \exists t_q \in \mathscr{T}_j^q \text{ such that } p - q\epsilon \in t_q,$$

so we can remove some open g from t while keeping that $\forall x \in [0, 1]$, $\Xi_j(x) \in \{2n, 2n + 1\}$.

Create Gaps

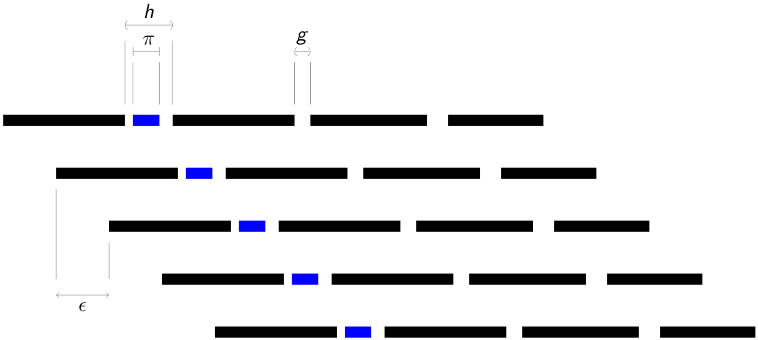

Figure: Sketch of creating a gap to our previous scenario.

Create Gaps

Assign $\psi_{j+1}(t_-) = \psi_j(t)$, and choose value for $\psi_{j+1}(t_+)$ so that:

- Maintain monotonicity
- Slope is bounded $\leq \frac{1}{2}$

Update $\mathscr{T}_j$ to $\mathscr{T}_{j+1}$ by replacing t with t_-, t_+.

Table of Contents

Implementation

- Implemented in Python, in serial (for now)
- Stores one system of towns $\mathscr{T}_j$ as an Interval Tree
- Extended precision

Results: Towns

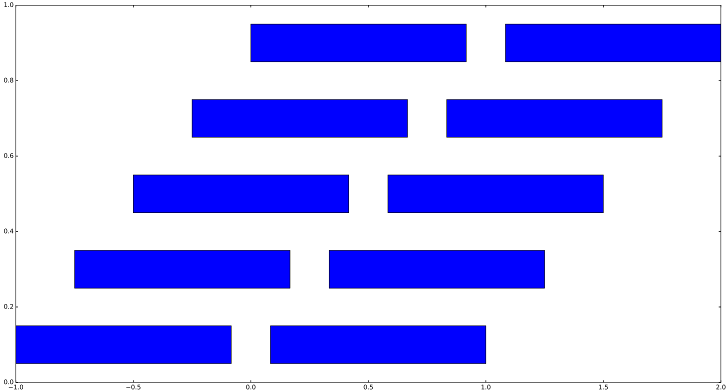

Figure: System of towns, after refinement $j = 1$

Results: Towns

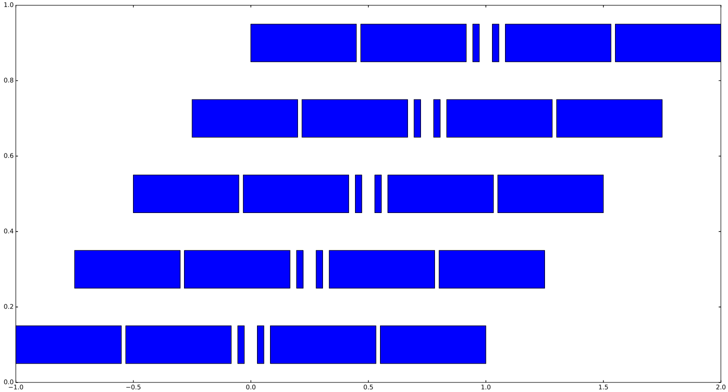

Figure: System of towns, after refinement $j = 2$

Results: Towns

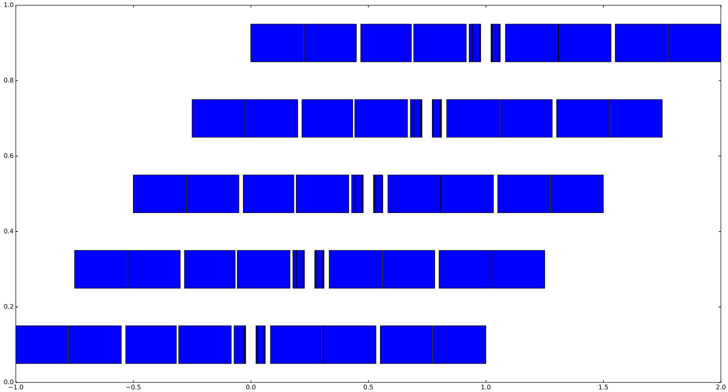

Figure: System of towns, after refinement $j = 3$

Results: ψ

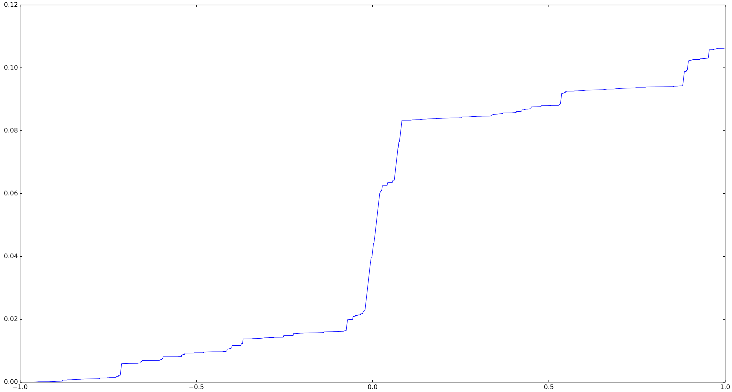

Figure: Function ψ generated after 11 iterations.

Table of Contents

Outer Function

Given f, need to construct outer function:

- Dependent on f
- Only need one (Lorenz 1966)
- If f is differentiable, needs to cancel out the non-differentiability of inner function
- Constructed iteratively by bounding oscillation and refining
- Implement in code...