Tropical Geometry, OMT and Neural Network

Na Lei\textsuperscript{1}

\textsuperscript{1}DUT-RU International School of Information Sciences and Engineering
Dalian University of Technology

Geometric Analysis Approach to AI
CMSA, Harvard
2019-01-22
Thanks

Thanks for the invitation.
This is a joint work with Professor David Xianfeng Gu, Zhongxuan Luo, Lianbao Jin, Hao Wang and so on.
Outline

1. Introduction to Tropical Geometry
2. Tropical Geometry for Neural Network
3. Optimal Transport Mapping
4. OMT by Neural Network
Introduction to Tropical Geometry
Tropical geometry is based on tropical semiring.

### Tropical semiring

1. Tropical Addition: $\oplus$;
2. Tropical Multiplication: $\odot$;
3. Tropical semiring: $\mathbb{T} := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot)$.

Tropical semiring is also known as the max-plus algebra.
Tropical Algebra

Tropical operators

For \( x, y \in \mathbb{R} \),
1. their tropical sum is \( x \oplus y := \max\{x, y\} \);
2. their tropical product is \( x \odot y := x + y \);
3. the tropical quotient of \( x \) over \( y \) is \( x \oslash y := x - y \).

\(-\infty \oplus x = x = 0 \odot x, \quad -\infty \odot x = -\infty.\)

\(-\infty\): the tropical additive identity, 0: the tropical multiplicative identity.

Associativity, Commutativity, Distributivity. Lack of additive inverse.

\( T := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot) \) is a semiring.
### Tropical operators

For $x, y \in \mathbb{R}$,

1. their tropical sum is $x \oplus y := \max\{x, y\};$
2. their tropical product is $x \odot y := x + y;$
3. the tropical quotient of $x$ over $y$ is $x \oslash y := x - y$.

$-\infty \oplus x = x = 0 \odot x, \quad -\infty \odot x = -\infty.$

$-\infty$: the tropical additive identity, 0: the tropical multiplicative identity.

Associativity, Commutativity, Distributivity. Lack of additive inverse.

$T := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot)$ is a semiring.
Tropical Algebra

Tropical operators

For \( x, y \in \mathbb{R} \),

1. their tropical sum is \( x \oplus y := \max\{x, y\} \);
2. their tropical product is \( x \odot y := x + y \);
3. the tropical quotient of \( x \) over \( y \) is \( x \oslash y := x - y \).

\[ -\infty \oplus x = x = 0 \odot x, \quad -\infty \odot x = -\infty. \]

\(-\infty\): the tropical additive identity, \(0\): the tropical multiplicative identity.

Associativity, Commutativity, Distributivity. Lack of additive inverse.

\( T := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot) \) is a semiring.
Tropical Algebra

Tropical operators

For $x, y \in \mathbb{R}$,

1. their tropical sum is $x \oplus y := \max\{x, y\}$;
2. their tropical product is $x \odot y := x + y$;
3. the tropical quotient of $x$ over $y$ is $x \oslash y := x - y$.

$-\infty \oplus x = x = 0 \odot x, \quad -\infty \odot x = -\infty$.

$-\infty$: the tropical additive identity, 0: the tropical multiplicative identity.

Associativity, Commutativity, Distributivity. Lack of additive inverse.

\[\mathbb{T} := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot)\] is a semiring.
Tropical power

Let $\mathbb{N} = \{ n \in \mathbb{Z} : n \geq 0 \}$. For any integer $a \in \mathbb{N}$, raising $x \in \mathbb{R}$ to the $a$–th power:

$$x \circ a := x \circ \cdots \circ x = a \cdot x$$

where the last $\cdot$ denotes standard product of real numbers. For any $a \in \mathbb{N}$,

$$-\infty \circ a := \begin{cases} -\infty & \text{if } a > 0, \\ 0 & \text{if } a = 0. \end{cases}$$

Every $x \in \mathbb{R}$ has a tropical multiplicative inverse given by its standard additive inverse:

$$x \circ -1 := -x.$$

Thus $\mathbb{T}$ is in fact a semifield.
**Tropical power**

Let \( \mathbb{N} = \{ n \in \mathbb{Z} : n \geq 0 \} \). For any integer \( a \in \mathbb{N} \), raising \( x \in \mathbb{R} \) to the \( a \)-th power:

\[
x^{\circ a} := x \circ \cdots \circ x = a \cdot x
\]

where the last \( \cdot \) denotes standard product of real numbers.

For any \( a \in \mathbb{N} \),

\[
-\infty^{\circ a} := \begin{cases} 
-\infty & \text{if } a > 0, \\
0 & \text{if } a = 0.
\end{cases}
\]

Every \( x \in \mathbb{R} \) has a tropical multiplicative inverse given by its standard additive inverse:

\[
x^{\circ -1} := -x.
\]

Thus \( \mathbb{T} \) is in fact a semifield.
One may therefore also raise $x \in \mathbb{R}$ to a negative power $a \in \mathbb{Z}$ by raising its tropical multiplicative inverse $-x$ to the positive power $-a$, i.e., $x \circ^a = (-x) \circ^{-a}$.

$-\infty$ does not have a tropical multiplicative inverse and $(-\infty) \circ^a$ is undefined for $a < 0$. 
One may therefore also raise \( x \in \mathbb{R} \) to a negative power \( a \in \mathbb{Z} \) by raising its tropical multiplicative inverse \( -x \) to the positive power \( -a \), i.e., \( x^{\odot a} = (-x)^{\odot -a} \).

\(-\infty\) does not have a tropical multiplicative inverse and \((-\infty)^{\odot a}\) is undefined for \( a < 0 \).
A tropical monomial in \(d\) variables \(x_1, \ldots, x_d\) is an expression of the form
\[
c x^\alpha = c \circ x_1^{a_1} \circ x_2^{a_2} \circ \cdots \circ x_d^{a_d}
\]
where \(c \in \mathbb{R} \cup \{-\infty\}, \ a_1, \ldots, a_d \in \mathbb{N}, \ \alpha = (a_1, \ldots, a_d) \in \mathbb{N}^d\) and \(x = (x_1, \ldots, x_d)\).

Note that \(x^\alpha = 0 \circ x^\alpha\) as 0 is the tropical multiplicative identity.

A tropical monomial is indeed a linear function, or equivalently, a hyperplane: \(a_1 x_1 + a_2 x_2 + \cdots + a_d x_d + c\).
A tropical monomial in $d$ variables $x_1, \ldots, x_d$ is an expression of the form

$$cx^\alpha = c \odot x_1^{\circ a_1} \odot x_2^{\circ a_2} \odot \cdots \odot x_d^{\circ a_d}$$

where $c \in \mathbb{R} \cup \{-\infty\}$, $a_1, \ldots, a_d \in \mathbb{N}$, $\alpha = (a_1, \ldots, a_d) \in \mathbb{N}^d$ and $x = (x_1, \ldots, x_d)$.

Note that $x^\alpha = 0 \odot x^\alpha$ as 0 is the tropical multiplicative identity.

A tropical monomial is indeed a linear function, or equivalently, a hyperplane: $a_1x_1 + a_2x_2 + \cdots + a_dx_d + c$. 
Tropical polynomial

A tropical polynomial \( f(x) = f(x_1, \ldots, x_d) \) is a finite tropical sum of tropical monomials

\[
f(x) = c_1 x^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r},
\]

where \( c_i \in \mathbb{R} \cup \{-\infty\} \) and \( \alpha_i = (a_{i1}, \ldots, a_{id}) \in \mathbb{N}^d \) for \( i = 1, \ldots, r \).

A monomial of a given multiindex appears at most once in the sum, i.e., \( \alpha_i \neq \alpha_j \) for any \( i \neq j \).

A tropical polynomial is indeed a piecewise linear function, or equivalently, an upper envelope of the \( r \) hyperplanes:

\[
\max_{i=1}^r \{ a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{id} x_d + c_i \}.
\]
Tropical Polynomial

A tropical polynomial \( f(x) = f(x_1, \ldots, x_d) \) is a finite tropical sum of tropical monomials

\[
f(x) = c_1 x_{\alpha_1} \oplus \cdots \oplus c_r x_{\alpha_r},
\]

where \( c_i \in \mathbb{R} \cup \{-\infty\} \) and \( \alpha_i = (a_{i1}, \ldots, a_{id}) \in \mathbb{N}^d \) for \( i = 1, \ldots, r \).

A monomial of a given multiindex appears at most once in the sum, i.e., \( \alpha_i \neq \alpha_j \) for any \( i \neq j \).

A tropical polynomial is indeed a piecewise linear function, or equivalently, an upper envelope of the \( r \) hyperplanes:

\[
\max_{i=1}^{r} \{ a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{id} x_d + c_i \}.
\]
Tropical Algebra

Tropical rational function

A tropical rational function is a standard difference, or, equivalently, a tropical quotient of two tropical polynomials $f(x)$ and $g(x)$:

$$f(x) \odot g(x) = f(x) - g(x).$$

The set of tropical polynomials $\mathbb{T}[x_1, \ldots, x_d]$ forms a semiring under the standard extension of $\oplus$ and $\otimes$ to tropical polynomials, and likewise the set of tropical rational functions $\mathbb{T}(x_1, \ldots, x_d)$ forms a semifield. We regard a tropical polynomial $f = f \odot 0$ as a special cases of a tropical rational function and thus

$$\mathbb{T}[x_1, \ldots, x_d] \subseteq \mathbb{T}(x_1, \ldots, x_d).$$
Tropical rational function

A tropical rational function is a standard difference, or, equivalently, a tropical quotient of two tropical polynomials $f(x)$ and $g(x)$:

$$f(x) \ominus g(x) = f(x) - g(x).$$

The set of tropical polynomials $\mathbb{T}[x_1, \ldots, x_d]$ forms a semiring under the standard extension of $\oplus$ and $\odot$ to tropical polynomials, and likewise the set of tropical rational functions $\mathbb{T}(x_1, \ldots, x_d)$ forms a semifield.

We regard a tropical polynomial $f = f \ominus 0$ as a special cases of a tropical rational function and thus

$$\mathbb{T}[x_1, \ldots, x_d] \subseteq \mathbb{T}(x_1, \ldots, x_d).$$
A tropical rational function is a standard difference, or, equivalently, a tropical quotient of two tropical polynomials $f(x)$ and $g(x)$:

$$f(x) \odot g(x) = f(x) - g(x).$$

The set of tropical polynomials $\mathbb{T}[x_1, \ldots, x_d]$ forms a semiring under the standard extension of $\oplus$ and $\odot$ to tropical polynomials, and likewise the set of tropical rational functions $\mathbb{T}(x_1, \ldots, x_d)$ forms a semifield. We regard a tropical polynomial $f = f \odot 0$ as a special cases of a tropical rational function and thus

$$\mathbb{T}[x_1, \ldots, x_d] \subseteq \mathbb{T}(x_1, \ldots, x_d).$$
The tropical hypersurface of \( f(x) = c_1 x^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r} \) is

\[
\mathcal{T}(f) := \{ x \in \mathbb{R}^d : c_i x^{\alpha_i} = c_j x^{\alpha_j} = f(x) \text{ for some } \alpha_i \neq \alpha_j \}.
\]
i.e., the set of points \( x \) at which the value of \( f \) at \( x \) is attained by two or more monomials in \( f \).

The tropical curve of

\[
1 \oplus x^2 \oplus 1 \oplus y^2 \oplus 2 \oplus xy \oplus 2 \oplus x \oplus 2 \oplus y \oplus 2.
\]
i.e.

\[
\max\{1 + 2x, 1 + 2y, 2 + x + y, 2 + x, 2 + y, 2\}.
\]
Tropical Geometry

Tropical hypersurface

The tropical hypersurface of $f(x) = c_1 x^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r}$ is

$$\mathcal{T}(f) := \{ x \in \mathbb{R}^d : c_i x^{\alpha_i} = c_j x^{\alpha_j} = f(x) \text{ for some } \alpha_i \neq \alpha_j \}.$$ 

i.e., the set of points $x$ at which the value of $f$ at $x$ is attained by two or more monomials in $f$.

The tropical curve of

$$1 \circledast x^2 \oplus 1 \circledast y^2 \oplus 2 \circledast xy \oplus 2 \circledast x \oplus 2 \circledast y \oplus 2.$$ 

i.e.

$$\max\{1 + 2x, 1 + 2y, 2 + x + y, 2 + x, 2 + y, 2\}.$$ 

The tropical hypersurfaces indeed induce a cell decomposition of $\mathbb{R}^d$. 

Na Lei (nalei@dlut.edu.cn) (DLUT)
The Newton polygon of a tropical polynomial $f(x) = c_1 x^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r}$ is the convex hull of $\alpha_1, \ldots, \alpha_d \in \mathbb{N}^d$, regarded as points in $\mathbb{R}^d$, is

$$\triangle(f) := \text{Conv}\{\alpha_i \in \mathbb{R}^d : c_i \neq -\infty, i = 1, \cdots, r\}.$$
Dual subdivision

Lift each $\alpha_i$ from $\mathbb{R}^d$ into $\mathbb{R}^{d+1}$ by appending $c_i$ as the last coordinate. Denote the convex hull of the lifted $\alpha_1, \ldots, \alpha_r$ as

$$\mathcal{P}(f) := \text{Conv}\{(\alpha_i, c_i) \in \mathbb{R}^d \times \mathbb{R} : i = 1, \ldots, r\}.$$ 

$\text{UF}(\mathcal{P}(f))$ denote the collection of upper faces in $\mathcal{P}(f)$ and $\pi : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ be the projection that drops the last coordinate. The dual subdivision determined by $f$ is then

$$\delta(f) := \{\pi(p) \subseteq \mathbb{R}^d : p \in \text{UF}(\mathcal{P}(f))\}$$

$\delta(f)$ forms a polyhedral complex with support $\triangle(f)$. 
Tropical Geometry

Upper envelope of polytope

Dual subdivision of Newton polygon
Dual subdivision

The tropical hypersurface $\mathcal{T}(f)$ is the $(d - 1)$-skeleton of the polyhedral complex dual to $\delta(f)$. This means that each vertex in $\delta(f)$ corresponds to one “cell” in $\mathbb{R}^d$ where the function $f$ is linear. Thus, the number of vertices in $\mathcal{P}(f)$ provides an upper bound on the number of linear regions of $f$.

The Newton Polygon and the tropical curve of

$$1 \circ x_1^2 + 1 \circ x_2^2 + 2 \circ x_1 \circ x_2 + 2 \circ x_1 + 2 \circ x_2 + 2.$$
A linear region of $F \in \text{Rat}(d, m)$ is a maximal connected subset of the domain on which $F$ is linear. The number of linear regions of $F$ is denoted $\mathcal{N}(F)$.

A tropical polynomial map $F \in \text{Pol}(d, m)$ has convex linear regions but a tropical rational map $F \in \text{Rat}(d, m)$ generally has nonconvex linear regions.
Neural Networks
Neural Networks

\[
\begin{align*}
\nu^{(l+1)} &= \sigma^{(l+1)} \circ \rho^{(l+1)} \circ \sigma^{(l)} \circ \rho^{(l)}(x) \\
\text{ith neuron in layer } l: &\quad \max \left\{ \sum_{j=1}^{n_l-1} a_{ij}^{(l)} y_j^{(l-1)} + b_i^{(l)} \right\} \\
\text{ith neuron in layer } l, \\&\quad \text{tropical formulation:} \\
&\quad \{ \left( \bigcirc_{j=1}^{n_l-1} (\nu_j^{(l-1)} a_{ij}^{(l)}) \right) \bigcirc b_i^{(l)} \bigoplus t_i^{(l)} \}
\end{align*}
\]
An $L$-layer feedforward neural network is a map $v: \mathbb{R}^d \rightarrow \mathbb{R}^p$ given by a composition of functions

$$v = \sigma^{(L)} \circ \rho^{(L)} \circ \sigma^{(L-1)} \circ \rho^{(L-1)} \circ \cdots \circ \sigma^{(1)} \circ \rho^{(1)}.$$ 

The preactivation functions $\rho^{(1)}, \ldots, \rho^{(L)}$ are affine transformations to be determined and the activation functions $\sigma^{(1)}, \ldots, \sigma^{(L)}$ are chosen and fixed in advance.
Neural Networks

Denote by $n_l$ the number of nodes of the $l$-th layer, $l = 1, \cdots, L - 1$. $n_0 := d$, $n_L := p$. The output from the $l$-th layer be denoted by

$$
\nu^{(l)} := \sigma^{(l)} \circ \rho^{(l)} \circ \sigma^{(l-1)} \circ \rho^{(l-1)} \circ \cdots \circ \sigma^{(1)} \circ \rho^{(1)}.
$$

The affine function $\rho^{(l)} : \mathbb{R}^{n_{l-1}} \to \mathbb{R}^{n_l}$ is given by a weight matrix $A^{(l)} = (a^{(l)}_{ij}) \in \mathbb{Z}^{n_l \times n_{l-1}}$ and a bias vector $b^{(l)} = (b^{(l)}_1, \cdots, b^{(l)}_{n_l})^T \in \mathbb{R}^{n_l}$:

$$
\rho^{(l)}(\nu^{(l-1)}) := A^{(l)} \nu^{(l-1)} + b^{(l)}.
$$
Neural Networks

Denote by $n_l$ the number of nodes of the $l$-th layer, $l = 1, \cdots, L - 1$. $n_0 := d$, $n_L := p$. The output from the $l$-th layer be denoted by

$$
\nu^{(l)} := \sigma^{(l)} \circ \rho^{(l)} \circ \sigma^{(l-1)} \circ \rho^{(l-1)} \circ \cdots \circ \sigma^{(1)} \circ \rho^{(1)}.
$$

The affine function $\rho^{(l)} : \mathbb{R}^{n_{l-1}} \rightarrow \mathbb{R}^{n_l}$ is given by a weight matrix $A^{(l)} = (a_{ij}^{(l)}) \in \mathbb{Z}^{n_l \times n_{l-1}}$ and a bias vector $b^{(l)} = (b_1^{(l)}, \cdots, b_{n_l}^{(l)})^T \in \mathbb{R}^{n_l}$:

$$
\rho^{(l)}(\nu^{(l-1)}) := A^{(l)} \nu^{(l-1)} + b^{(l)}.
$$
Neural Networks

Three Assumptions

(a) The weight matrices \( A^{(1)}, \ldots, A^{(L)} \) are integer-valued;
(b) The bias vectors \( b^{(1)}, \ldots, b^{(L)} \) are real-valued;
(c) The activation functions \( \sigma^{(1)}, \ldots, \sigma^{(L)} \) take the form

\[
\sigma^{(l)}(x) := \max\{x, t^{(l)}\}
\]

where \( t^{(l)} \in (\mathbb{R} \cup \{-\infty\})^{n_l} \) is called a threshold vector.

- Real weights can be approximated arbitrarily closely by rational weights;
- One may then ‘clear denominators’ in these rational weights by multiplying them by the least common multiple of their denominators to obtain integer weights;
- Scaling all weights and biases by the same positive constant has no bearing on the workings of a neural network.
Three Assumptions

(a) The weight matrices $A^{(1)}, \ldots, A^{(L)}$ are integer-valued;
(b) The bias vectors $b^{(1)}, \ldots, b^{(L)}$ are real-valued;
(c) The activation functions $\sigma^{(l)}, \ldots, \sigma^{(L)}$ take the form

$$\sigma^{(l)}(x) := \max\{x, t^{(l)}\}$$

where $t^{(l)} \in (\mathbb{R} \cup \{-\infty\})^{n_l}$ is called a threshold vector.

- real weights can be approximated arbitrarily closely by rational weights;
- one may then ‘clear denominators’ in these rational weights by multiplying them by the least common multiple of their denominators to obtain integer weights;
- scaling all weights and biases by the same positive constant has no bearing on the workings of a neural network.
Tropical Algebra of Neural Networks
Consider the output from one layer in the neural network

\[ v(x) = \max\{Ax + b, t\}, \]

Decompose \( A \) as a difference of two nonnegative matrices
\( A = A_+ - A_- \) with entries

\[ a_{ij}^+ := \max\{a_{ij}, 0\}, \quad a_{ij}^- = \max\{-a_{ij}, 0\}. \]

Since

\[ \max\{Ax + b, t\} = \max\{A_+ x + b, A_- x + t\} - A_- x, \]

every coordinate of one-layer neural network is a difference of two tropical polynomials.
Proposition

If the nodes of the \( l \)-th layer are given by tropical rational functions,

\[
\nu^{(l)}(x) = F^{(l)}(x) \odot G^{(l)}(x) = F^{(l)}(x) - G^{(l)}(x),
\]

then the outputs of the preactivation and of the \((l+1)\)-th layer are given by tropical rational functions

\[
\rho^{(l+1)} \circ \nu^{(l)}(x) = H^{(l+1)}(x) - G^{(l+1)}(x),
\]

\[
\nu^{(l+1)}(x) = \sigma \circ \rho^{(l+1)} \circ \nu^{(l)}(x) = F^{(l+1)}(x) - G^{(l+1)}(x)
\]

respectively, where

\[
F^{(l+1)}(x) = \max\{H^{(l+1)}(x), G^{(l+1)}(x) + t\},
\]

\[
G^{(l+1)}(x) = A_+ G^{(l)}(x) + A_- F^{(l)}(x),
\]

\[
H^{(l+1)}(x) = A_+ F^{(l)}(x) + A_- G^{(l)}(x) + b.
\]
Tropical characterization of neural networks

A feedforward neural network under assumptions (a)-(c) is a function $\nu: \mathbb{R}^d \rightarrow \mathbb{R}^p$ whose coordinates are tropical rational functions of the input, i.e.,

$$\nu(x) = F(x) \circ G(x) = F(x) - G(x),$$

where $F$ and $G$ are tropical polynomial maps. Thus $\nu$ is a tropical rational map.

Corollary

Setting $t^{(1)} = \ldots = t^{(L-1)} = 0$ and $t^{(L)} = -\infty$. Let $\nu: \mathbb{R}^d \rightarrow \mathbb{R}$ be an ReLU activated feedforward neural network with integer weights and linear output. Then $\nu$ is a tropical rational function.
(i) Let $\nu : \mathbb{R}^d \to \mathbb{R}$. Then $\nu$ is a tropical rational function if and only if $\nu$ is a feedforward neural network satisfying assumptions (a)-(c).

(ii) A tropical rational function $f \odot g$ can be represented as an $L$-layer neural network, with

$$L \leq \max\{\lceil \log_2 r_f \rceil, \lceil \log_2 r_g \rceil \} + 2,$$

where $r_f$ and $r_g$ are the number of monomials in the tropical polynomials $f$ and $g$ respectively.
Continuous piecewise linear function

Let $\nu : \mathbb{R}^d \to \mathbb{R}$. Then $\nu$ is a continuous piecewise linear function with integer coefficients if and only if $\nu$ is a tropical rational function.

Equivalence

(i) Tropical rational functions,
(ii) Continuous piecewise linear functions with integer coefficients,
(iii) Neural networks satisfying assumptions (a)-(c).
Continuous piecewise linear function

Let $\nu : \mathbb{R}^d \to \mathbb{R}$. Then $\nu$ is a continuous piecewise linear function with integer coefficients if and only if $\nu$ is a tropical rational function.

Equivalence

(i) Tropical rational functions,
(ii) Continuous piecewise linear functions with integer coefficients,
(iii) Neural networks satisfying assumptions (a)-(c).
Optimal Transport Mapping
Earth movement cost.
Optimal Mass Transportation

Problem Setting

Find the best scheme of transporting one mass distribution \((\mu, U)\) to another one \((\nu, V)\) such that the total cost is minimized, where \(U, V\) are two bounded domains in \(\mathbb{R}^n\), such that

\[
\int_U \mu(x)\,dx = \int_V \nu(y)\,dy,
\]

\(0 \leq \mu \in L^1(U)\) and \(0 \leq \nu \in L^1(V)\) are density functions.
For a transport scheme \( s \) (a mapping from \( U \) to \( V \))

\[
s : x \in U \rightarrow y \in V,
\]

the total cost is

\[
C(s) = \int_{U} \mu(x)c(x, s(x)) \, dx
\]

where \( c(x, y) \) is the cost function.
Cost Function $c(x, y)$

The cost of moving a unit mass from point $x$ to point $y$.

$Monge(1781) : c(x, y) = |x - y|$

This is the natural cost function. Other cost functions include

\[
\begin{align*}
  c(x, y) &= |x - y|^p, p \neq 0 \\
  c(x, y) &= -\log |x - y| \\
  c(x, y) &= \sqrt{\varepsilon + |x - y|^2}, \varepsilon > 0
\end{align*}
\]

Any function can be cost function. It can be negative.
Problem

Is there an optimal mapping $T : U \rightarrow V$ such that the total cost $\mathcal{C}$ is minimized,

$$\mathcal{C}(T) = \inf \{ \mathcal{C}(s) : s \in \mathcal{S} \}$$

where $\mathcal{S}$ is the set of all measure preserving mappings, namely $s : U \rightarrow V$ satisfies

$$\int_{s^{-1}(E)} \mu(x) dx = \int_E v(y) dy, \forall \text{ Borel set } E \subset V$$
Solutions

Three categories:

1. Discrete category: both \((\mu, U)\) and \((\nu, V)\) are discrete,
2. Semi-continuous category: \((\mu, U)\) is continuous, \((\nu, V)\) is discrete,
3. Continuous category: both \((\mu, U)\) and \((\nu, V)\) are continuous.
Kantorovich’s Approach

Both \((\mu, U)\) and \((\nu, V)\) are discrete. \(\mu\) and \(\nu\) are Dirac measures. 

\((\mu, U)\) is represented as

\[
\{(\mu_1, p_1), (\mu_2, p_2), \ldots, (\mu_m, p_m)\},
\]

\((\nu, V)\) is

\[
\{(\nu_1, q_1), (\nu_2, q_2), \ldots, (\nu_n, q_n)\}.
\]

A transportation plan \(f: \{p_i\} \rightarrow \{q_j\}, f = \{f_{ij}\}\), \(f_{ij}\) means how much mass is moved from \((\mu_i, p_i)\) to \((\nu_j, q_j)\), \(i \leq m, j \leq n\). The optimal mass transportation plan is:

\[
\min_f f_{ij} c(p_i, q_j)
\]

with constraints:

\[
\sum_{j=1}^{n} f_{ij} = \mu_i, \sum_{i=1}^{m} f_{ij} = \nu_j.
\]

Optimizing a linear energy on a convex set, solvable by linear programming method.
Kantorovich’s Approach

Kantorovich won Nobel’s prize in economics.

\[
\min \sum_{ij} f_{ij} c(p_i, p_j),
\]

such that

\[
\sum_j f_{ij} = \mu_i, \sum_i f_{ij} = \nu_j.
\]

There are \(mn\) unknowns in total. The complexity is quite high.
Brenier’s Approach

Theorem (Brenier)

If $\mu, \nu > 0$ and $U$ is convex, and the cost function is quadratic distance,

$$c(x, y) = |x - y|^2$$

then there exists a convex function $f : U \to \mathbb{R}$ unique upto a constant, such that the unique optimal transportation map is given by the gradient map

$$T : x \to \nabla f(x).$$
Brenier’s Approach

Continuous Category: In smooth case, the Brenier potential $f : U \to \mathbb{R}$ satisfies the Monge-Ampere equation

$$\text{det} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right) = \frac{\mu(x)}{\nu(\nabla f(x))},$$

and $\nabla f : U \to V$ minimizes the quadratic cost

$$\min_f \int_U |x - \nabla f(x)|^2 \, dx.$$
Given a compact convex domain $U$ in $\mathbb{R}^n$ and $p_1, \cdots, p_k$ in $\mathbb{R}^n$ and $A_1, \cdots, A_k > 0$, find a transport map $T : U \to \{p_1, \cdots, p_k\}$ with $\text{vol}(T^{-1}(p_i)) = A_i$, so that $T$ minimizes the transport cost

$$\int_U |x - T(x)|^2 dx.$$
Theorem (Aurenhammer-Hoffmann-Aronov 1998)

Given a compact convex domain $U$ in $\mathbb{R}^n$ and $p_1, \cdots, p_k$ in $\mathbb{R}^n$ and $A_1, \cdots, A_k > 0$, $\sum_i A_i = \text{vol}(U)$, there exists a unique power diagram

$$U = \bigcup_{i=1}^{k} W_i,$$

$\text{vol}(W_i) = A_i$, the map $T : W_i \mapsto p_i$ minimizes the transport cost

$$\int_U |x - T(x)|^2 dx.$$
Voronoi Decomposition
Voronoi Diagram

Given $p_1, \cdots, p_k$ in $\mathbb{R}^n$, the Voronoi cell $W_i$ at $p_i$ is

$$W_i = \{x | |x - p_i|^2 \leq |x - p_j|^2, \forall j\}.$$
Power Distance

Given $p_i$ associated with a sphere $(p_i, r_i)$ the power distance from $q \in \mathbb{R}^n$ to $p_i$ is

$$pow(p_i, q) = |p_i - q|^2 - r_i^2.$$
Power Diagram

Given \( p_1, \cdots, p_k \) in \( \mathbb{R}^n \) and power weights \( h_1, \cdots, h_k \), the power Voronoi cell \( W_i \) at \( p_i \) is

\[
W_i = \{ x \mid |x - p_i|^2 + h_i \leq |x - p_j|^2 + h_j, \forall j \}.
\]
Lemma

Suppose \( f(x) = \max\{\langle x, p_i \rangle + h_i \} \) is a piecewise linear convex function, then its gradient map induces a power diagram,

\[
W_i = \{ x | \nabla f = p_i \}.
\]

Proof.

\[
\langle x, p_i \rangle + h_i \geq \langle x, p_j \rangle + h_j
\]
is equivalent to

\[
|x - p_i|^2 - 2h_i - |p_i|^2 \leq |x - p_j|^2 - 2h_j - |p_j|^2.
\]
A Piecewise Linear convex function

\[ f(x) := \max\{\langle x, p_i \rangle + h_i | i = 1, \ldots, k \} \]

produces a convex cell decomposition \( W_i \) of \( \mathbb{R}^n \):

\[ W_i = \{ x | \langle x, p_i \rangle + h_i \geq \langle x, p_j \rangle + h_j, \forall j \} \]

Namely, \( W_i = \{ x | \nabla f(x) = p_i \} \).
Alexandrov Theorem

**Theorem (Alexandrov 1950)**

Given $\Omega$ compact convex domain in $\mathbb{R}^n$, $p_1, \cdots, p_k$ distinct in $\mathbb{R}^n$, $A_1, \cdots, A_k > 0$, such that $\sum A_i = \text{Vol}(\Omega)$, there exists PL convex function

$$f(x) := \max\{\langle x, p_i \rangle + h_i | i = 1, \cdots, k\}$$

unique up to translation such that

$$\text{Vol}(W_i) = \text{Vol}(\{x | \nabla f(x) = p_i\}) = A_i.$$

Alexandrov’s proof is topological, not variational. It has been open for years to find a constructive proof.
Variational Proof

Theorem (Gu-Luo-Sun-Yau 2013)

\( \Omega \) is a compact convex domain in \( \mathbb{R}^n \), \( y_1, \cdots, y_k \) distinct in \( \mathbb{R}^n \), \( \mu \) a positive continuous measure on \( \Omega \). For any \( \nu_1, \cdots, \nu_k > 0 \) with \( \sum \nu_i = \mu(\Omega) \), there exists a vector \((h_1, \cdots, h_k)\) so that

\[
     u(x) = \max \{ \langle x, p_i \rangle + h_i \}
\]

satisfies \( \mu(W_i \cap \Omega) = \nu_i \), where \( W_i = \{x | \nabla f(x) = p_i\} \). Furthermore, \( h \) is the maximum point of the convex function

\[
     E(h) = \sum_{i=1}^{k} \nu_i h_i - \int_{0}^{h} \sum_{i=1}^{k} w_i(\eta) d\eta_i,
\]

where \( w_i(\eta) = \mu(W_i(\eta) \cap \Omega) \) is the \( \mu \)-volume of the cell.
One can define a cylinder through $\partial \Omega$, the cylinder is truncated by the $xy$-plane and the convex polyhedron. The energy term $\int h \sum w_i(\eta) d\eta_i$ equals to the volume of the truncated cylinder.
Definition (Alexandrov Potential)

The concave energy is

\[ E(h_1, h_2, \ldots, h_k) = \sum_{i=1}^{k} v_i h_i - \int_0^h \sum_{j=1}^{k} w_j(\eta) d\eta_j, \]

Geometrically, the energy is the volume beneath the parabola.
The gradient of the Alexanrov potential is the differences between the target measure and the current measure of each cell

\[ \nabla E(h_1, h_2, \cdots, h_k) = (\nu_1 - w_1, \nu_2 - w_2, \cdots, \nu_k - w_k) \]
The essential computation of OMT is to compute a piecewise linear function.

Equivalence

(i) Tropical rational functions,
(ii) Continuous piecewise linear function: OMT,
(iii) Neural networks.
OMT by Neural Networks
## OT vs. Tropical Geometry vs. NN

<table>
<thead>
<tr>
<th>Optimal Transport</th>
<th>Tropical Geometry</th>
<th>Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample data $p_i$</td>
<td>tropical power $p_i$</td>
<td>weight matrix</td>
</tr>
<tr>
<td>hyperplane $\langle x, p_i \rangle + h_i$</td>
<td>tropical monomial $h_i \odot x^{p_i}$</td>
<td>one neuron</td>
</tr>
<tr>
<td>upper envelope $\max_i {\langle x, p_i \rangle + h_i}$</td>
<td>tropical polynomial</td>
<td>neural network</td>
</tr>
<tr>
<td>power diagram $\oplus_i h_i \odot x^{p_i}$</td>
<td>tropical hypersurface</td>
<td>activated path</td>
</tr>
<tr>
<td>power weights $h_i$</td>
<td>unknowns $h_i$</td>
<td>bias</td>
</tr>
</tbody>
</table>

### Diagrams

- **Tropical Polynomial**
  - $\pi_i(h)$
  - $\nabla u_h$
  - $W_i(h)$
  - $\text{proj}$

- **Upper Envelope of Polytope**
  - $\pi_i^*$
  - $\text{proj}^*$

- **Tropical Hypersurfaces**
  - $\text{proj}^*$

- **Subdivision of Newton Polygon**
  - $y_i$
Merits

OMT map can be represented as a tropical polynomial, and further realized using neural network,

- the weights are fixed, only train the biases;
- the number of parameters equal to that of the training samples;
- the method is independent of the dimension, general for arbitrary dimension;
- convex optimization, fast convergence, easy to train.
OMT-NN Generative Model

(a) training data

(b) generated data
Conclusion

- The fundamental concepts of tropical geometry;
- The equivalence relation between tropical rational function and ReLU DNN;
- The tropical geometric representation of OMT;
- The realization of OMT using neural network.
Thank you!

For more information, please email to nalei@dlut.edu.cn.