Last time:

Intro & Background. & Monge-Kantorovich

To day:

- Background:
  - Problem setting
  - Functional analysis complement.
    * Legendre transform
    * Fenchel duality (w. proof)
    * Hahn-Banach theorem
- General Duality proof: only for compactness & C_b.
  * Fenchel duality & Riesz-Markov Theorem
  * Approximation
- Special cases: distance cost function
  * Kantorovich-Rubinstein theorem
  * Connection to total variation distance.

✡ Problem Setting:

▷ (X, μ), (Y, ν) 2 probability spaces, $c: X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ measurable.

Kantorovich mass transportation problem:

$$\underset{\pi \in \Pi(\mu,\nu)}{\arg\min} \int_{X \times Y} c(x,y)\, d\pi(x,y)$$

- rmk 1: $\Pi(\mu,\nu)$ is nonempty & convex.
- rmk 2: From now we only consider Borel prob. sp.:

  $(X, \mathcal{B}_X, \mu)$ where $\mathcal{B}_X := \sigma\left(\{O : O \text{ open in } X\}\right)$
- rmk 3: $\pi \in \Pi(\mu,\nu) \iff \forall (\varphi, \psi) \in L^1(\mu) \times L^1(\nu) = L^\infty(\mu) \times L^\infty(\nu)$

$$(*) \quad \int_{X \times Y} [\varphi(x) + \psi(y)]\, d\pi = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu =: J(\varphi, \psi)$$

  - rmk 3.1:

    If X, Y are Polish | separable complete metric sp, which the theorem concerns.

    $(*)$ holds for $(\varphi, \psi) \in C_b(X) \times C_b(Y)$ is sufficient.
  - rmk 3.2:

    If X, Y are Polish & locally compact | $\forall$ point has a cpt nbhd.

    $(*)$ holds for $(\varphi, \psi) \in C_0(X) \times C_0(Y)$ is sufficient.

✡ General Duality:

- Thm: [Kantorovich duality]
  - (X,μ), (Y,ν) Polish. Cost function $c: X\times Y \to \overline{\mathbb{R}}_+$ lower semi-continuous | $\forall x \in X$, $c(x) \le \liminf_{y\to x} c(y)$
  - $\forall \pi \in \mathcal{P}(X\times Y)$, $(\varphi,\psi) \in L^1(\mu)\times L^1(\nu)$, define

$$I[\pi] = \int_{X\times Y} c(x,y)\, d\pi(x,y) \qquad J(\varphi,\psi) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu.$$

  - Define $\Pi(\mu,\nu)$ the set of all Borel coupling of μ,ν.
  - Define $\Phi_c = \{(\varphi,\psi): X\times Y \to \overline{\mathbb{R}}$ meas : $\varphi(x)+\psi(y) \le c(x,y)$ for $d\mu$-almost all $x\in X$, $d\nu$-almost all $y \in Y\} \cap \left(L^1(\mu)\times L^1(\nu)\right)$

  Then

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi)$$

- rmk 4: The infimum, may infinite, can be attained.
- rmk 5: The supreme does not change if $\Phi_c = \Phi_c \cap \left(C_b(X)\times C_b(Y)\right)$.

▷ Before proof:

- Legendre (- Fenchel) Transform

  E a normed vector space. $\Theta : E \to \mathbb{R} \cup \{+\infty\}$ convex.

  Then the Legendre transform is the function $\Theta^* : E^* \to \mathbb{R}$ by

$$\Theta^*(z^*) = \sup_{z \in E} \left[\langle z^*, z \rangle - \Theta(z) \right].$$

- rmk 6:

6.1.1. Definition. *Let X be a normed space. The space $X^* := \mathcal{L}(X, \mathbb{R})$ (or $X^* := \mathcal{L}(X, \mathbb{C})$ in the complex case) of all continuous linear functionals on the space X is called the dual (or topological dual) to the space X. The space X' of all linear functions on X is called the algebraic dual.*

- Fenchel (- Rockafellar) Duality

  - E (separable) normed vector sp.
  - $\Theta, \Xi : E \to \mathbb{R} \cup \{+\infty\}$. Θ^*, Ξ^* Legendre trans of Θ and Ξ.
  - Assume $\exists\ x_0 \in E$ s.t. $\Theta(z_0), \Xi(z_0) < +\infty$. Θ cts at z_0.

Then

$$\inf_E [\Theta + \Xi] = \max_{z^* \in E^*} \left[-\Theta^*(-z^*) - \Xi^*(z^*) \right]$$

$$\max_{z^* \in E^*} = \sup_{z^* \in E^*}$$

Proof: WTS:

$$RHS = \sup_{z^*\in E^*}\left[-\Theta^*(-z^*) - \Xi^*(z^*)\right]$$

$$= \sup_{z^*\in E^*}\left\{-\sup_{x\in E}\left[\langle -z^*, x\rangle - \Theta(x)\right] - \sup_{y\in E}\left[\langle z^*, y\rangle - \Xi(y)\right]\right\}$$

$$= \sup_{z^*\in E^*}\left\{\inf_{x\in E}\left[\langle z^*, x\rangle + \Theta(x)\right] + \inf_{y\in E}\left[\langle z^*, -y\rangle + \Xi(y)\right]\right\}$$

$$= \sup_{z^*\in E^*}\ \inf_{x,y\in E}\left\{\left[\langle z^*, x-y\rangle + \Theta(x) + \Xi(y)\right]\right\} \qquad (1)$$

$$= LHS = \inf_{x\in E}\left[\Theta(x) + \Xi(x)\right] \qquad (2)$$

① $(1) \leq (2)$

$x = y$

$$\Rightarrow (1) \leq \sup_{z^*}\ \inf_{x}\left[\Theta(x) + \Xi(x)\right] = \inf_{x}\left[\Theta(x) + \Xi(x)\right] = (2)$$

$\Theta(z_0)$, $\Xi(z_0)$ finite $\Rightarrow m := \inf_{x}\left[\Theta(x) + \Xi(x)\right] \leq \Theta(z_0) + \Xi(z_0) \leq +\infty$

② $(1) \geq (2)$

Consider $C := \{(x, \lambda) \in E\times\mathbb{R} : \lambda > \Theta(x)\}$ ← hypergraph of convex fct

and $C' := \{(y, \mu) \in E\times\mathbb{R} : \mu \leq m - \Xi(y)\}$ ← subgraph of concave fct

- both convex.

- $\forall x$, $\Theta(x) + \Xi(x) \geq m \Rightarrow \Theta(x) \geq m - \Xi(x)$

  $\Rightarrow C \cap C' = \phi$ (disjoint)

- $(z_0, \Theta(z_0)+1) \in C \Rightarrow \mathrm{Int}(C) \neq \phi$

By Hahn - Banach separation theorem :

Theorem[19] — Let A and B be non-empty convex subsets of a real locally convex topological vector space X. If $\operatorname{Int} A \neq \varnothing$ and $B \cap \operatorname{Int} A = \varnothing$ then there exists a continuous linear functional f on X such that $\sup f(A) \leq \inf f(B)$ and $f(a) < \inf f(B)$ for all $a \in \operatorname{Int} A$ (such an f is necessarily non-zero).

$\exists$ a non-trivial linear form $\ell \in (E \times \mathbb{R})^*$ s.t.

$$\inf \langle \ell, c \rangle = \inf_{c \in \operatorname{Int}(C)} \langle \ell, c \rangle \geqslant \sup_{c' \in C'} \langle \ell, c' \rangle . \qquad (*)$$

Write ℓ as $(w^*, \alpha) \neq \vec{0}$, $\forall (x, \lambda) \in C$ and $(y, \mu) \in C'$,

$(*) \Leftrightarrow$ $\langle w^*, x \rangle + \alpha \lambda \geqslant \langle w^*, y \rangle + \alpha \mu$, where $\alpha > 0$ (Think about subsp. $\vec{0} \times \mathbb{R}$)

Normalize w^* s.t. $z^* = \dfrac{w^*}{\alpha}$ and we have

$$(*) \Leftrightarrow \langle z^*, x \rangle + \lambda \geqslant \langle z^*, y \rangle + \mu$$

$$\Big\downarrow \lambda > \Theta(x) \qquad \qquad \Big\downarrow \mu \leq m - \Xi(y)$$

$$\langle z^*, x \rangle + \Theta(x) \geqslant \langle z^*, y \rangle + m - \Xi(y)$$

$$\Rightarrow \langle z^*, x - y \rangle + \Theta(x) + \Xi(y) \geqslant m = LHS$$

$$\Rightarrow \inf_{x,y} [\quad -\|- \quad] \geqslant LHS = (2)$$

$$\Rightarrow \quad (1) = \sup_{z^*} \inf_{x,y} [-\|-] \geqslant LHS = (2)$$

$\therefore$ (1) = (2) ■

▷ Proof of Kantorovich Duality (Using Fenchel's Duality)

- We mainly consider step 1:
  $\mathcal{X}, \mathcal{Y}$ both compact, $c \in (C_b(\mathcal{X}\times\mathcal{Y}) =: E, \|\cdot\|_\infty)$

Recall Fenchel Duality

(i) E (separable) normed vector sp. (Satisfied!)

(ii) $\Theta, \Xi : E \to \mathbb{R}\cup\{+\infty\}$. Θ^*, Ξ^* Legendre trans of Θ and Ξ.

(iii) Assume $\exists\ z_0 \in E$ s.t. $\Theta(z_0), \Xi(z_0) < +\infty$. Θ cts at z_0.

Then $\inf_E [\Theta + \Xi] = \max_{z^*\in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)]$
$= \sup_{z^*\in E^*}$

- Two convex functions:

$$\Theta : u \in C_b(\mathcal{X},\mathcal{Y}) \longmapsto \begin{cases} 0, & \text{if } u(x,y) \geq -c(x,y) \\ +\infty, & \text{o.w.} \end{cases}$$

$$\Xi : u \in C_b(\mathcal{X},\mathcal{Y}) \longmapsto \begin{cases} \int_{\mathcal{X}} \varphi\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu, & \text{if } u(x,y) = \varphi(x) + \psi(y) \ (\cdots \text{ a.s.}) \\ +\infty & \text{, o.w.} \end{cases}$$

(ii) satisfied.

- $z_0(x,y) \equiv 1 \geq 0 \Rightarrow \Theta(z_0) = 0 < \infty$ and Θ cts at $z_0 \Rightarrow$ (iii) Satisfied.
  $z_0 \equiv \underset{\frac{1}{2}}{\underset{\parallel}{\varphi}} + \underset{\frac{1}{2}}{\underset{\parallel}{\psi}}$

Hence: $\inf_E [\Theta + \Xi] = \sup_{z^*\in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)]$

$$= \inf_E \left\{ \left(\int_{\mathcal{X}} \varphi\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu\right) \mathbb{1}_{\{\varphi(x)+\psi(y) = u(x,y) \geq -c(x,y)\}} + (+\infty)\, \mathbb{1}_{\{\cdots c\}} \right\}$$

$$= \inf_E \left\{ \int_{\mathcal{X}} \varphi\, d\mu + \int_{\mathcal{Y}} d\nu : \varphi(x) + \psi(y) \geq -c(x,y) \right\}$$

$$= -\sup_{\Phi_c} J(\varphi, \psi)$$

Now it remains to show $\sup_{z^* \in E^*} \left[-\Theta^*(-z^*) - \Xi^*(z^*) \right] = -\inf_{\Pi(\mu,\nu)} I(\pi)$.

$$\boxed{\Theta^*(z^*) = \sup_{z \in E} \left[\langle z^*, z \rangle - \Theta(z) \right].}$$

Question: What is z^*?
How to characterize z^*?

Riesz-Markov Theorem:

Theorem. Let X be a locally compact Hausdorff space. For any continuous linear functional ψ on $C_0(X)$, there is a unique *regular* countably additive complex Borel measure μ on X such that

$$\forall f \in C_0(X): \qquad \psi(f) = \int_X f(x)\, d\mu(x).$$

The norm of ψ as a linear functional is the total variation of μ, that is

$$\|\psi\| = |\mu|(X).$$

Finally, ψ is positive if and only if the measure μ is non-negative.

Here, we adopt Radon measure | Inner regular
locally finite

So, $\forall u^*$, $\exists$ unique Radon measure π s.t.

$$\langle u^*, u \rangle = u^*(u) = \int_{X \times Y} u(x,y)\, d\pi(x,y)$$

Okay. now we have

$$\Theta : u \in C_b(X,Y) \longmapsto \begin{cases} 0 & \text{if } u(x,y) \geq -c(x,y) \\ +\infty & \text{o.w.} \end{cases}$$

$$\Rightarrow \Theta^*(-\pi) = \sup_{u \in C_b(X \times Y)} \Big\{ -\int u(x,y)\, d\pi - \underbrace{\Theta(u)}_{\text{not be infinite}} \Big\}$$

$$= \sup_{u \in C_b(X \times Y)} \Big\{ \int -u(x,y)\, d\pi : -u(x,y) \leq c(x,y) \Big\}$$

$$= \begin{cases} \int c\, d\pi, & \pi \text{ nonnegative measure} \\ +\infty, & \text{o.w.} \end{cases}$$

$$\Xi : u \in C_b(X,Y) \longmapsto \begin{cases} \int_X \varphi\, d\mu + \int_Y \psi\, d\nu, & \text{if } u(x,y) = \varphi(x) + \psi(y) \ (\cdots \text{a.s.}) \\ +\infty & \text{, o.w.} \end{cases}$$

$$\Rightarrow \Xi^*(\pi) = \sup_u \Big\{ \int u(x,y)\, d\pi - \Xi(u) \Big\}$$

$$= \sup_u \Big\{ \int [\varphi(x) + \psi(y)]\, d\pi - \Big(\int_X \varphi\, d\mu + \int_Y \psi\, d\nu\Big) \Big\}$$

$$= \begin{cases} 0, & \text{if } \int_{X \times Y} [\varphi(x) + \psi(y)]\, d\pi = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu \\ +\infty, & \text{o.w.} \end{cases}$$

$$\Updownarrow$$

$$(\varphi, \psi) \in \Phi_c$$

$$\sup_{z^* \in E^*} \left[-\Theta^*(-z^*) - \Xi^*(z^*) \right] = \sup_\pi \left[-\int c\, d\pi\ \mathbb{1}_{\{\pi \text{ nonnegative measure}\}} - \infty\, \mathbb{1}_{\{\ldots c\}} \right]$$

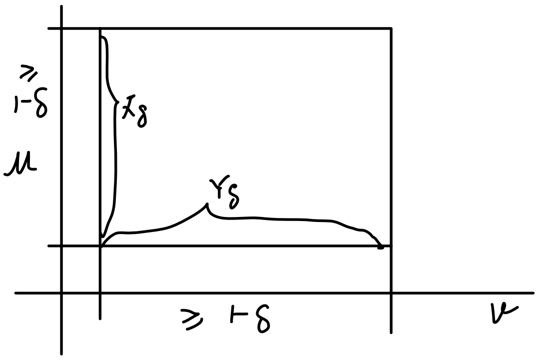
$$\int [\varphi(x) + \psi(y)]\, d\pi = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu$$

$$= -\inf_{\pi \in \Pi(\mu,\nu)} \int c\, d\pi = -\inf_{\pi \in \Pi(\mu,\nu)} I(\pi)$$

Q.E.D.

- Step 2: $\begin{cases} c \in C_b \\ \text{No assumption for } X, Y \text{ (May not compact)} \end{cases}$

Tightness $\Pi(\mu, \nu) \to$ compact subset $\to$ approximation (page 28 – 31)



$\delta \to 0$

- Step 3: No assumption for X, Y

No assumption for c (May be L^1 but not C_b)

$c = \sup c_n$, c_n non-decreasing seq of nonneg, unif cts cost fcts.

$I_n[\pi] = \int_{X\times Y} c_n \, d\pi$, $\sup_{(\varphi,\psi)\in\Phi_{c_n}} J(\varphi,\psi)$.

step 2: $\inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi] = \sup_{(\varphi,\psi)\in\Phi_{c_n}} J(\varphi,\psi)$

$\begin{cases} \sup_{(\varphi,\psi)\in C_n} J(\varphi,\psi) \le \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi) \\ \underbrace{\inf_{(\varphi,\psi)\in\Phi_{c_n}} I[\pi] = \sup_n \inf_{\pi\in\Phi_c} I_n[\pi]}_{} \end{cases}$

$\inf_{\Pi(\mu,\nu)} I[\pi] \le \sup_{\Phi_c} J(\varphi,\psi)$

[$\Pi(\mu,\nu)$ tight & Prokhorov's Theorem
MCT]

- Step 4: Infimum can be attained

✡ Distance cost function

- $c(x,y)$ may not be a metric.
- When the cost function is a metric $d(x,y)$, then there is more structure in the Kantorovich Duality principle
- Kantorovich – Rubinstein Theorem
  - $X = Y$ be a Polish space
  - $d: X \times X \to \mathbb{R}$ a lower semi-continuous metric on X.

  Let $T_d(\mu,\nu) = \inf\limits_{\pi \in \Pi(\mu,\nu)} \int_{X\times X} d(x,y)\, d\pi$ the cost of OT.

  Let $Lip(X)$ denote the space of all Lipschitz functions on X.

  Define $\|\varphi\|_{Lip} = \sup\limits_{x \neq y} \dfrac{|\varphi(x) - \varphi(y)|}{d(x,y)}$.

  Then

$$T_d(\mu,\nu) = \sup\left\{ \int_X \varphi \, d(\mu-\nu) ;\ \varphi \in L^1(|\mu-\nu|),\ \|\varphi\|_{Lip} \le 1 \right\}$$

$$|\varphi(x) - \varphi(y)| \le d(x,y)$$

- rmk 7: Let $P_1(X)$ be the space of those probability measures μ such that $\int d(x_0, x)\, d\mu(x) < +\infty$ for some (and thus any) x_0, and let $M_1(X)$ be the vector space generated by $P_1(X)$. On $M_1(X)$ we can define the norm

$$(1.20) \qquad \|\sigma\|_{KR} = \sup\left\{ \int_X \varphi \, d\sigma; \quad \varphi \in L^1(d|\sigma|), \quad \|\varphi\|_{\mathrm{Lip}} \le 1 \right\}.$$

Then the Kantorovich-Rubinstein theorem states that $\mathcal{T}_d(\mu,\nu) = \|\mu - \nu\|_{KR}$ for all probability measures μ, ν in $P_1(X)$.

When d is bounded, $\mathcal{T}_d$ and $\|\cdot\|_{KR}$ are well-defined on $P(X)$ and $M(X)$ respectively; then $\mathcal{T}_d$ is sometimes called the "bounded Lipschitz distance", and $\|\cdot\|_{KR}$ the "bounded Lipschitz norm".

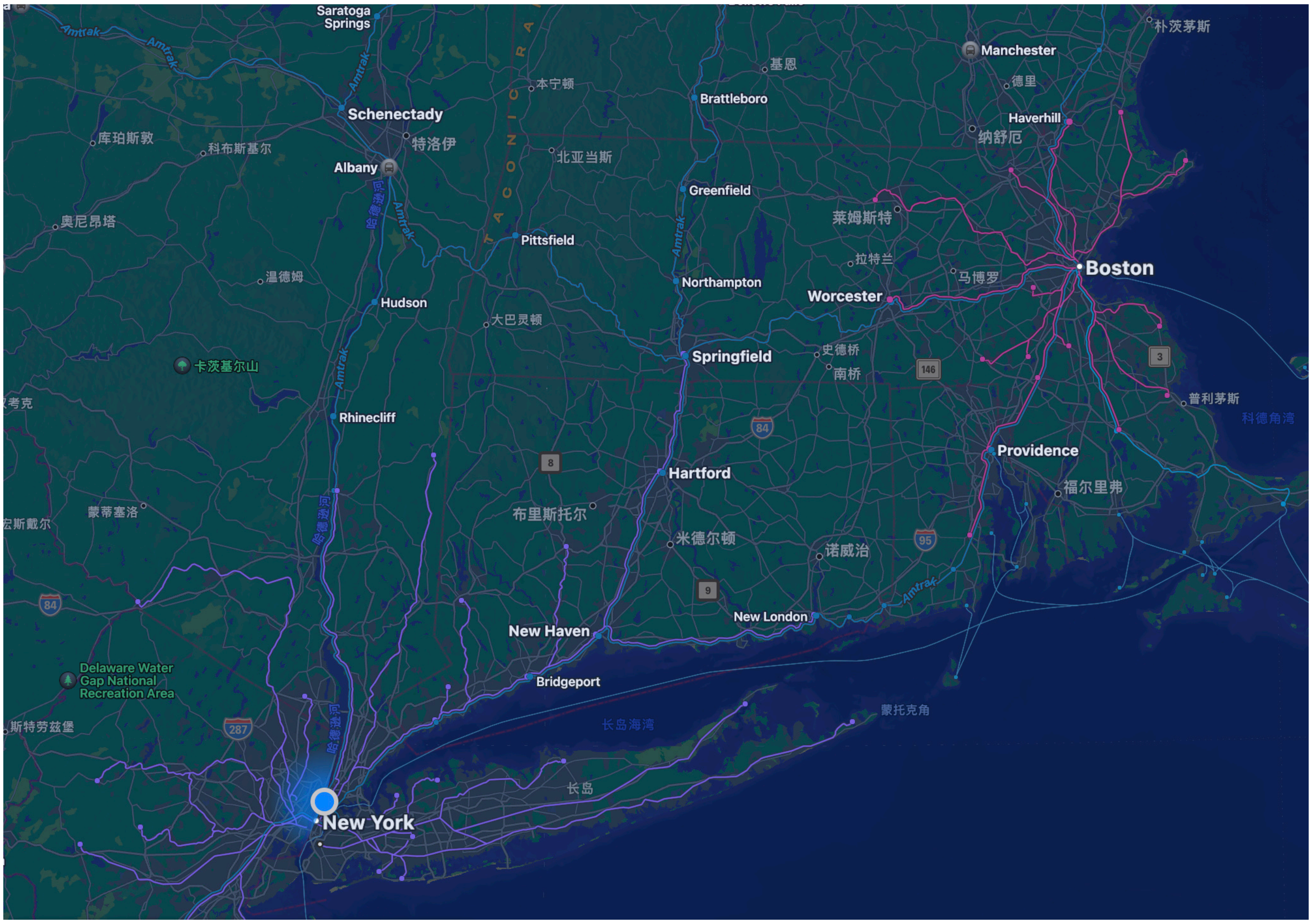
- Invariance of K-R distance
  - $X = Y$ Polish, d semi-cts metric
  - μ, ν, σ Borel non-negative finite measures on X.

  Then $T_d(\mu,\nu) = T_d(\mu+\sigma, \nu+\sigma)$

▷ Example: Total variation distance.

- X polish sp.
- cost function $c(x,y) = \mathbb{1}_{x \neq y}$
- $\mu, \nu \in P(X)$

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi(\{x \neq y\}) = \sup_{0 \le f \le 1} \int_X f \, d(\mu - \nu)$$


Saratoga Springs
朴茨茅斯
Manchester
基恩
本宁顿
德里
Schenectady
Brattleboro
Haverhill
库珀斯敦
科布斯基尔
特洛伊
纳舒厄
北亚当斯
Albany
Greenfield
奥尼昂塔
莱姆斯特
Pittsfield
拉特兰
Boston
温德姆
马博罗
Northampton
Hudson
Worcester
大巴灵顿
卡茨基尔山
Springfield
史德桥
南桥
Rhinecliff
普利茅斯
科德角湾
Providence
Hartford
蒙蒂塞洛
布里斯托尔
福尔里弗
米德尔顿
诺威治
New London
New Haven
Delaware Water Gap National Recreation Area
Bridgeport
蒙托克角
斯特劳兹堡
长岛海湾
长岛
New York
Amtrak