清華大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

Banker Online Mirror Descent

— A Universal Approach for Delayed Online Bandit Learning

Jiatai Huang *
Tsinghua University
hjt18@mails.tsinghua.edu.cn

Yan Dai *
Tsinghua University
yan-dai20@mails.tsinghua.edu.cn

Longbo Huang
Tsinghua University
longbohuang@tsinghua.edu.cn

[*]: Equal contribution.

Motivative Setting: Delayed Adversarial MAB

Motivative Setting: Delayed Adversarial MAB

- Delays $(d_1, d_2, \ldots, d_T)$ are chosen before-hand, but are kept unknown to the agent at all time

Motivative Setting: Delayed Adversarial MAB

- Delays $(d_1, d_2, \ldots, d_T)$ are chosen before-hand, but are kept unknown to the agent at all time
- Loss vectors $l_1, l_2, \ldots, l_T$ are adversarial chosen, but all entries are $[0,1]$-bounded (i.e., $l_t \in [0,1]^A$)

Motivative Setting: Delayed Adversarial MAB

- Delays $(d_1, d_2, \dots, d_T)$ are chosen before-hand, but are kept unknown to the agent at all time
- Loss vectors $l_1, l_2, \dots, l_T$ are adversarial chosen, but all entries are $[0,1]$-bounded (i.e., $l_t \in [0,1]^A$)
- Agent picks action A_t at each round $t = 1,2, \dots, T$, but only observes (t, l_{t,A_t}) at the end of round $t + d_t$

Motivative Setting: Delayed Adversarial MAB

- Delays $(d_1, d_2, \dots, d_T)$ are chosen before-hand, but are kept unknown to the agent at all time
- Loss vectors $l_1, l_2, \dots, l_T$ are adversarial chosen, but all entries are $[0,1]$-bounded (i.e., $l_t \in [0,1]^A$)
- Agent picks action A_t at each round $t = 1,2, \dots, T$, but only observes (t, l_{t,A_t}) at the end of round $t + d_t$

- **Optimal regret achieved by Zimmert et al. (2020):**

$$O\left(\sqrt{KT} + \sqrt{D \log K}\right).$$

Motivation of Our Work

Motivation of Our Work

- Delay model easily generalize to other problems

Motivation of Our Work

- Delay model easily generalize to other problems
  - Linear bandits
  - Combinatorial bandits
  - …

Motivation of Our Work

- Delay model easily generalize to other problems
  - Linear bandits
  - Combinatorial bandits
  - …
- Mostly studied on MABs **(Bistritz et al., 2019; Thune et al., 2019; Zimmert et al., 2020)**.

Motivation of Our Work

- Delay model easily generalize to other problems
  - Linear bandits
  - Combinatorial bandits
  - …
- Mostly studied on MABs **(Bistritz et al., 2019; Thune et al., 2019; Zimmert et al., 2020)**.
  - $O(\sqrt{KT} + \sqrt{D \log K})$ optimal regret **already achieved**

Motivation of Our Work

- Delay model easily generalize to other problems
  - Linear bandits
  - Combinatorial bandits
  - …
- Mostly studied on MABs **(Bistritz et al., 2019; Thune et al., 2019; Zimmert et al., 2020)**.
  - $O(\sqrt{KT} + \sqrt{D \log K})$ optimal regret **already achieved**
  - But… crucially depend on negative-entropy regularizer
  - Also task specific — not generalize to other problems

Motivation of Our Work

- Delay model easily generalize to other problems
  - Linear bandits
  - Combinatorial bandits
  - …
- Mostly studied on MABs **(Bistritz et al., 2019; Thune et al., 2019; Zimmert et al., 2020)**.
  - $O(\sqrt{KT} + \sqrt{D \log K})$ optimal regret **already achieved**
  - But… crucially depend on negative-entropy regularizer
  - Also task specific — not generalize to other problems
- Want **a universal approach** to handle delays robustly!

Classical Framework: OMD

Classical Framework: OMD

- Online Mirror Descent (OMD)

Classical Framework: OMD

- Online Mirror Descent (OMD)
  - Solves many online learning problems

Classical Framework: OMD

- Online Mirror Descent (OMD)
  - Solves many online learning problems
  - …… and their bandit-feedback versions

Classical Framework: OMD

- Online Mirror Descent (OMD)
  - Solves many online learning problems
  - …… and their bandit-feedback versions
  - ………… and their adversarial-loss versions

Classical Framework: OMD

- Online Mirror Descent (OMD)
  - Solves many online learning problems
  - …… and their bandit-feedback versions
  - ………… and their adversarial-loss versions
  - OMD Algorithm ≈ Regularizer + Step-sizes:

$$x_{t+1} = \arg\min_{x \in A} \left(\eta \langle \tilde{l}_t, x \rangle + D_\Psi(x, x_t) \right), \qquad \forall t.$$

Classical Framework: OMD

- Online Mirror Descent (OMD)
  - Solves many online learning problems
  - …… and their bandit-feedback versions
  - ………… and their adversarial-loss versions
  - OMD Algorithm ≈ Regularizer + Step-sizes:

$$x_{t+1} = \arg\min_{x \in A} \left(\eta \langle \tilde{l}_t, x \rangle + \mathrm{D}_\Psi(x, x_t) \right), \qquad \forall t.$$

  - "Greedily pick an action w.r.t. estimated loss, while keeping close to the last step"

Classical Framework: OMD

- Online Mirror Descent (OMD)
  - Solves many online learning problems
  - …… and their bandit-feedback versions
  - ………… and their adversarial-loss versions
  - OMD Algorithm ≈ Regularizer + Step-sizes:

$$x_{t+1} = \arg\min_{x \in A} \left(\eta \langle \tilde{l}_t, x \rangle + \mathrm{D}_\Psi(x, x_t) \right), \qquad \forall t.$$

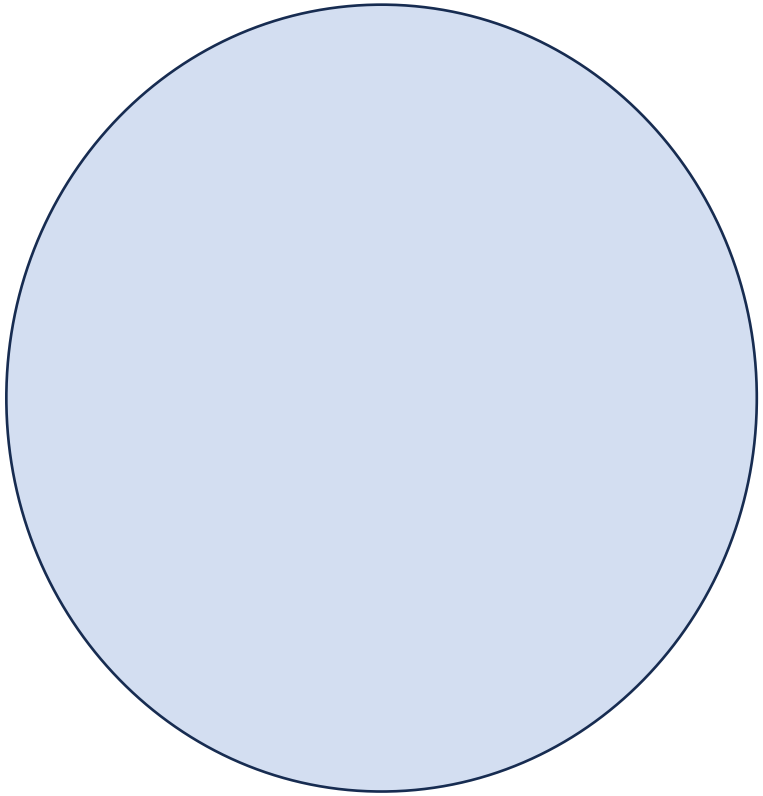
  - "Greedily pick an action w.r.t. estimated loss, while keeping close to the last step"

- Sadly, vanilla OMD cannot handle delays

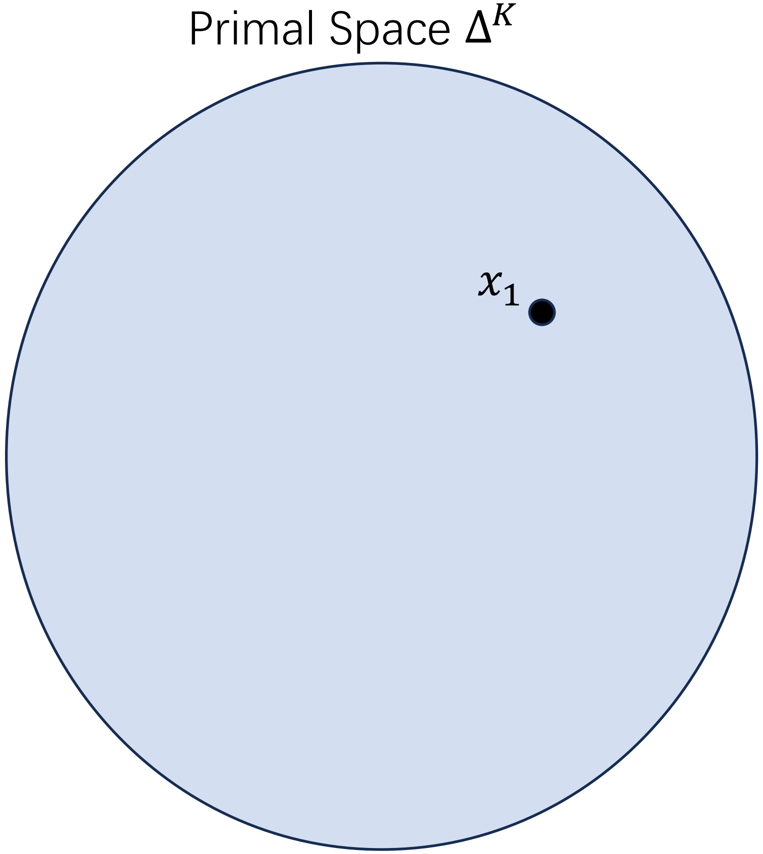
Vanilla OMD

Vanilla OMD

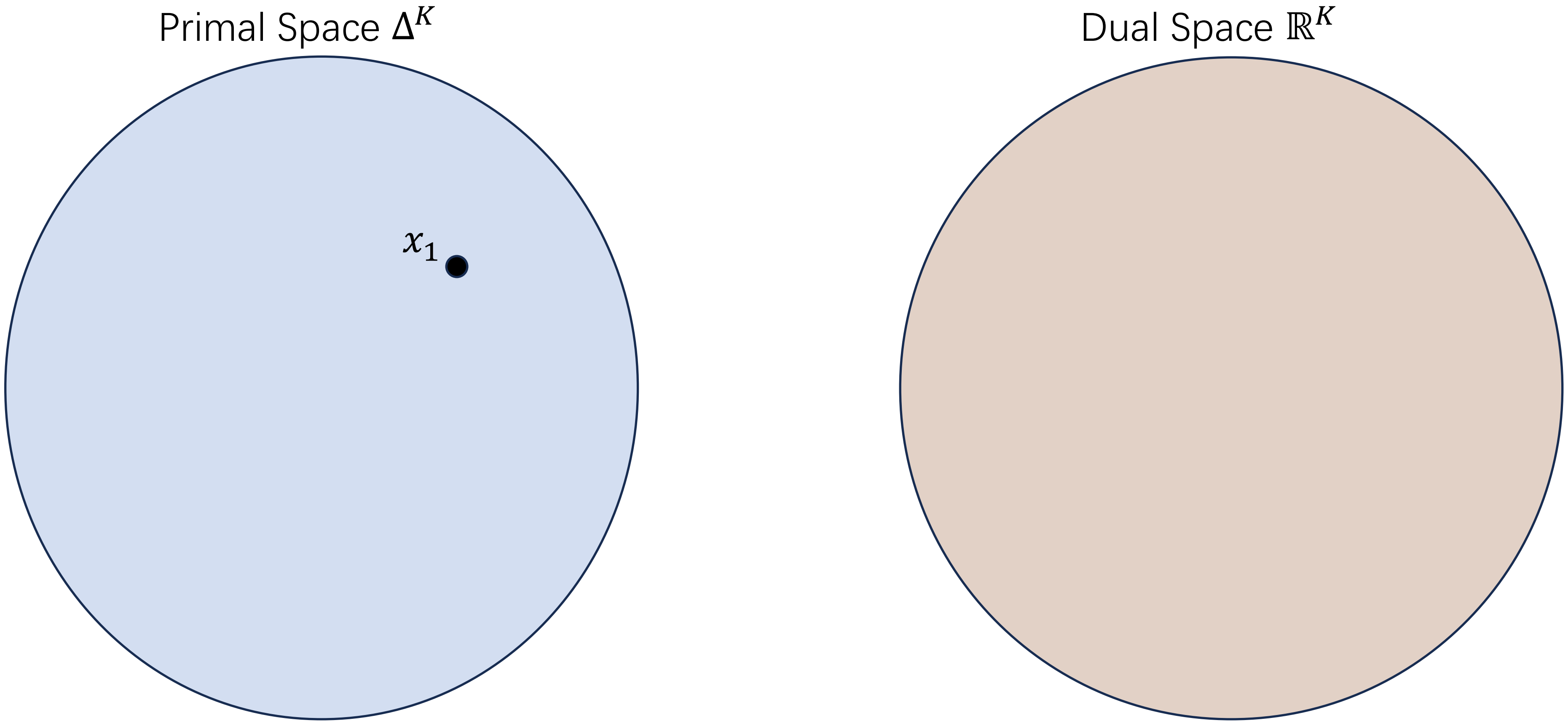
Primal Space Δ^K



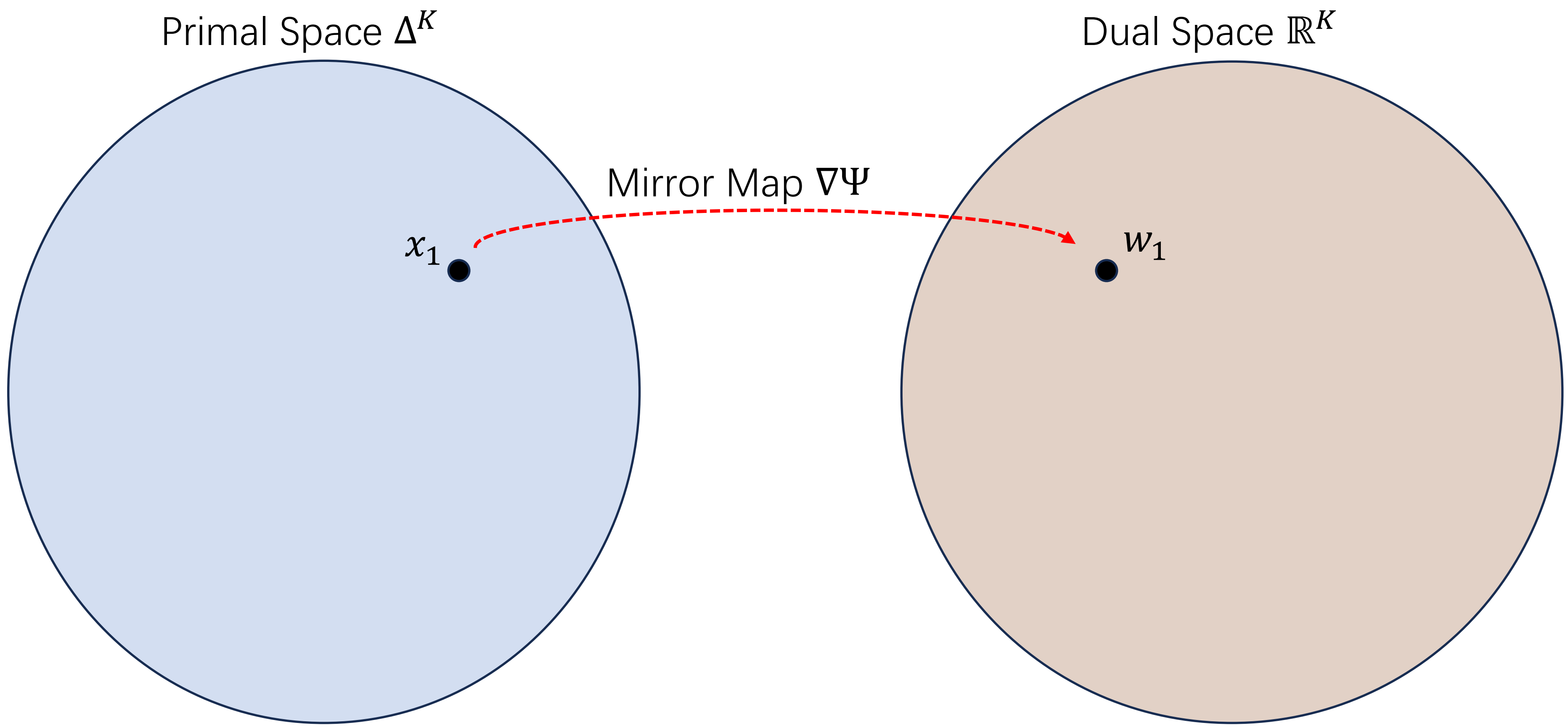
Vanilla OMD



Vanilla OMD

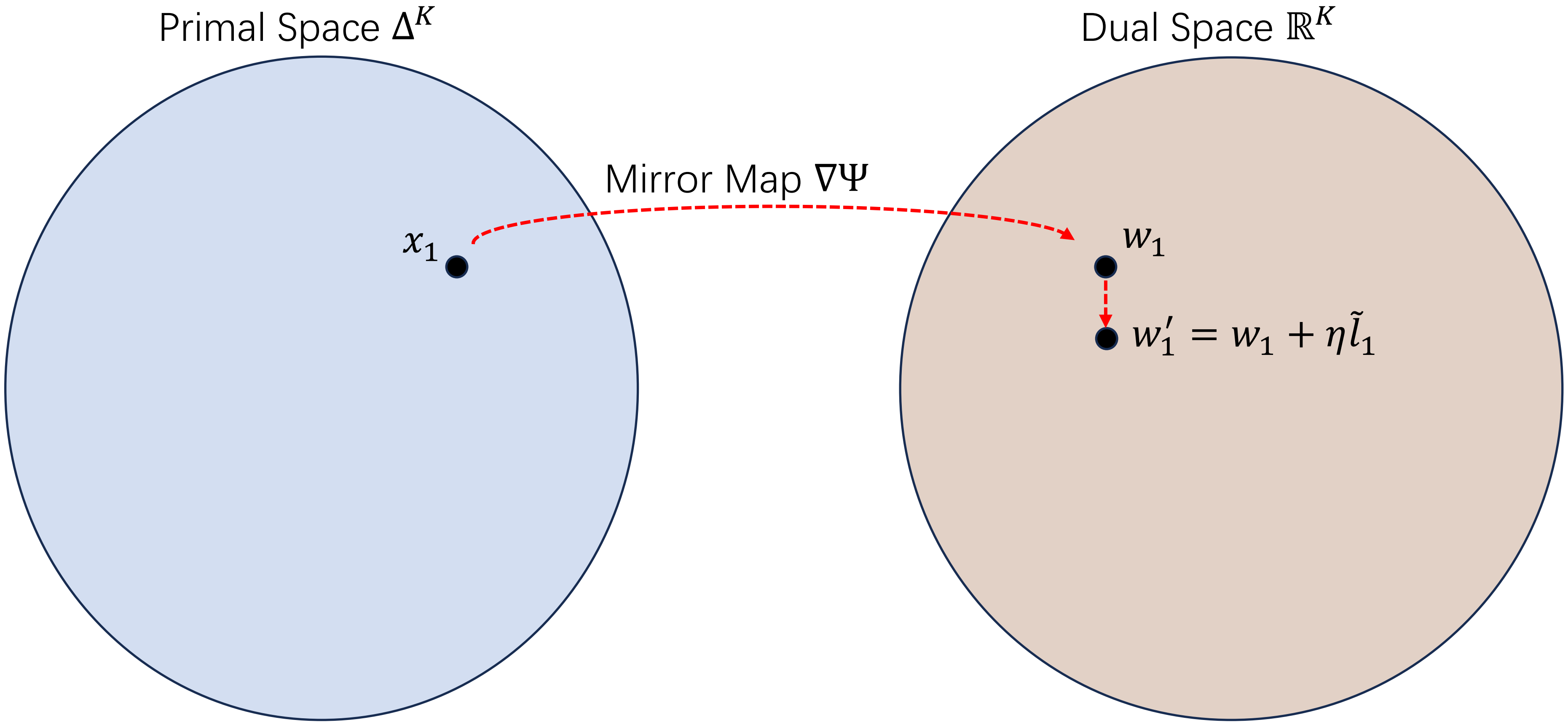


Vanilla OMD



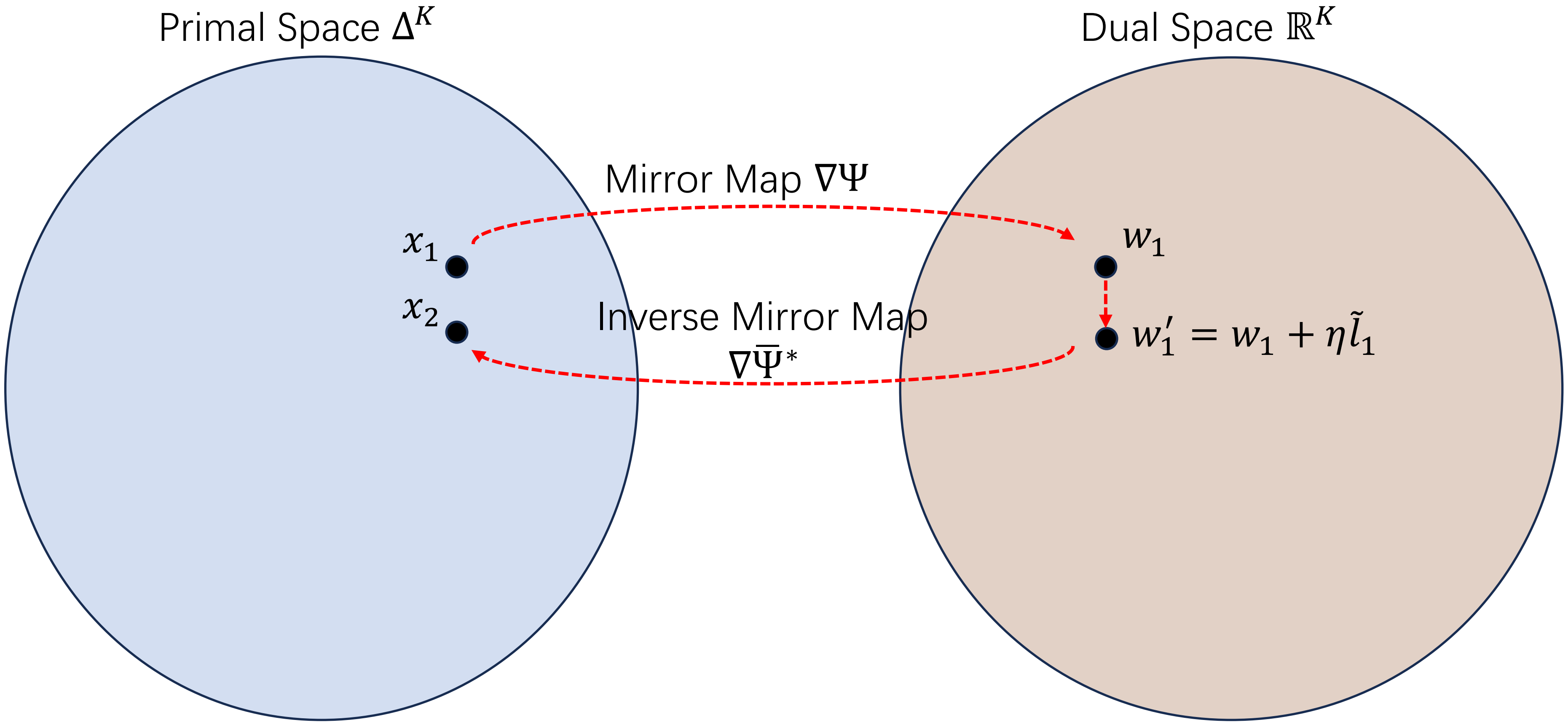
Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
x_1
w_1

Vanilla OMD



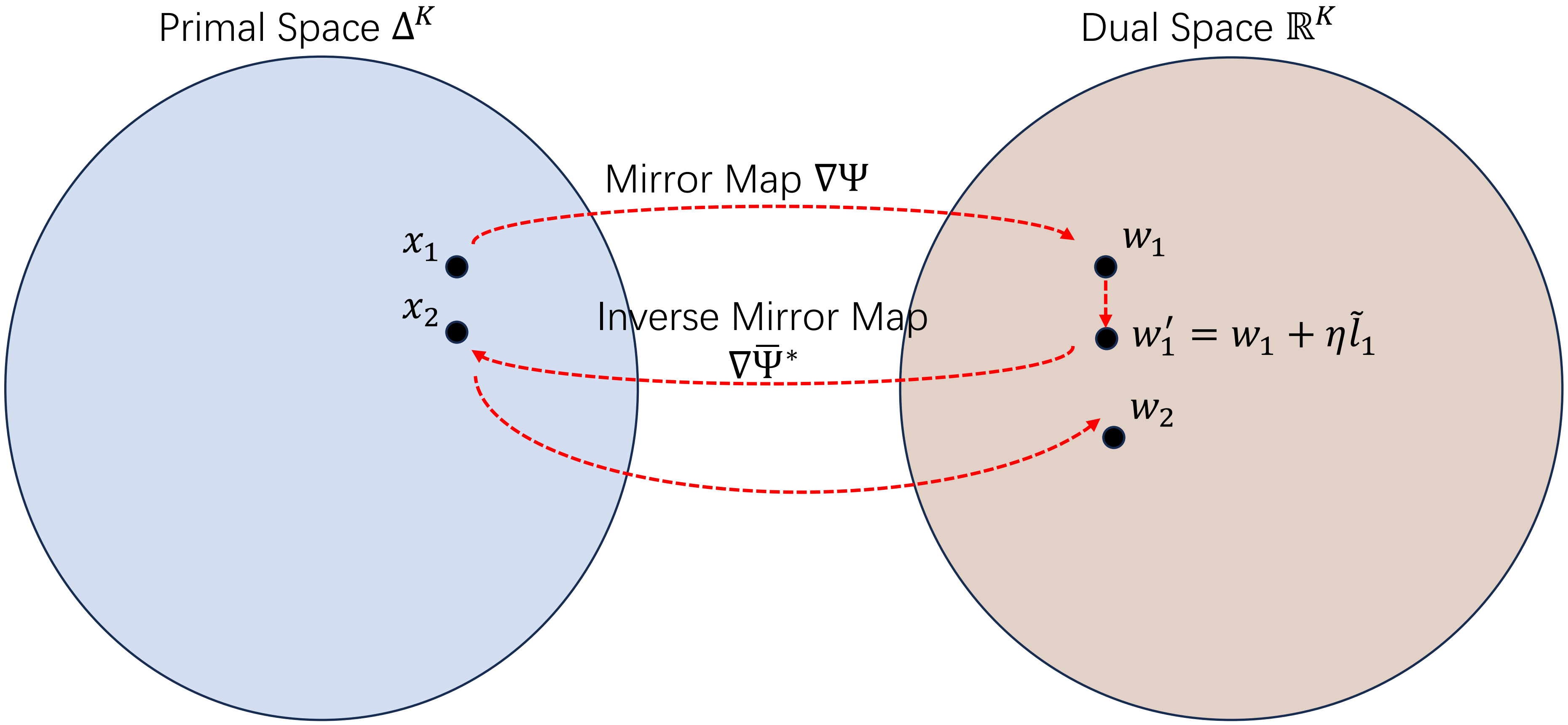
Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
x_1
w_1
$w_1' = w_1 + \eta\tilde{l}_1$

Vanilla OMD



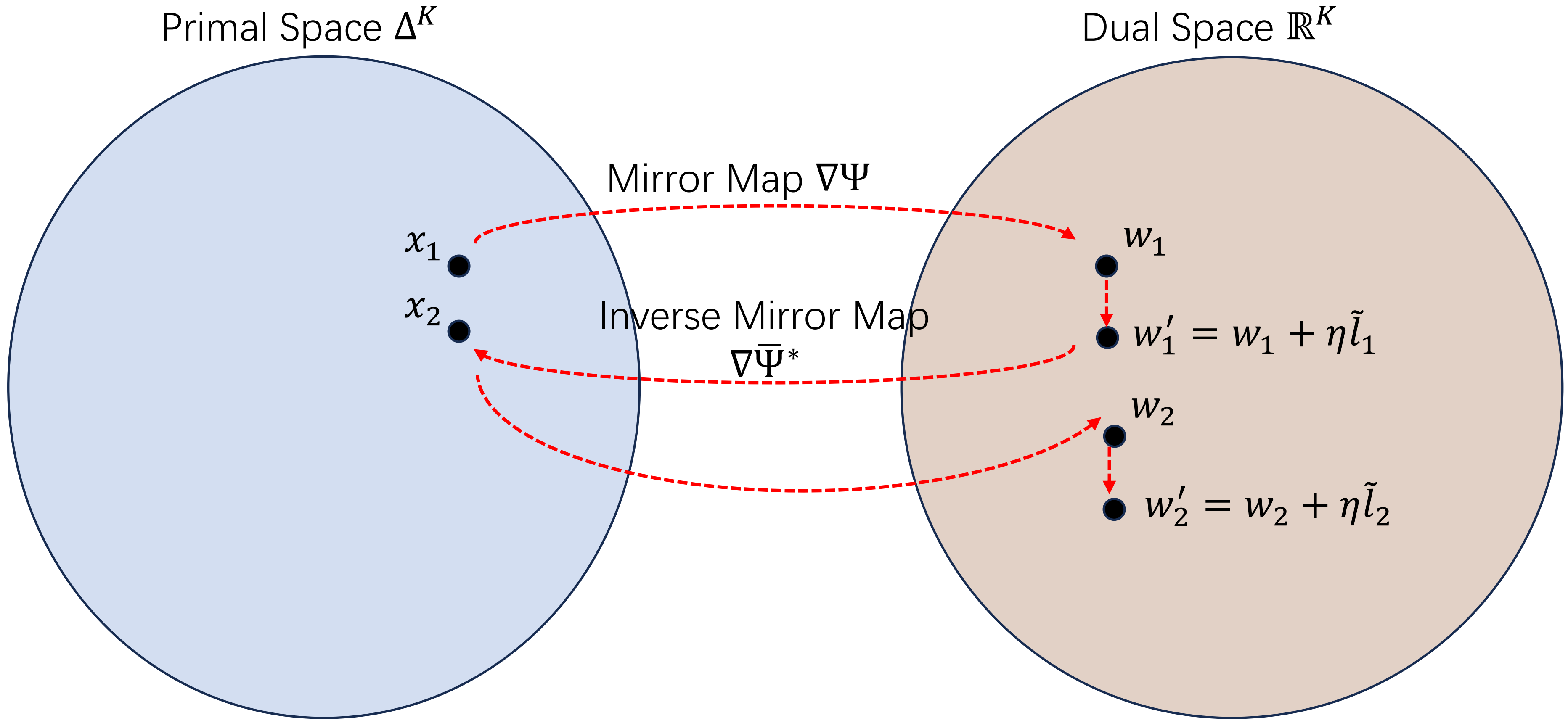
Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
x_1
x_2
w_1
Inverse Mirror Map
$\nabla\bar{\Psi}^*$
$w_1' = w_1 + \eta\tilde{l}_1$

Vanilla OMD



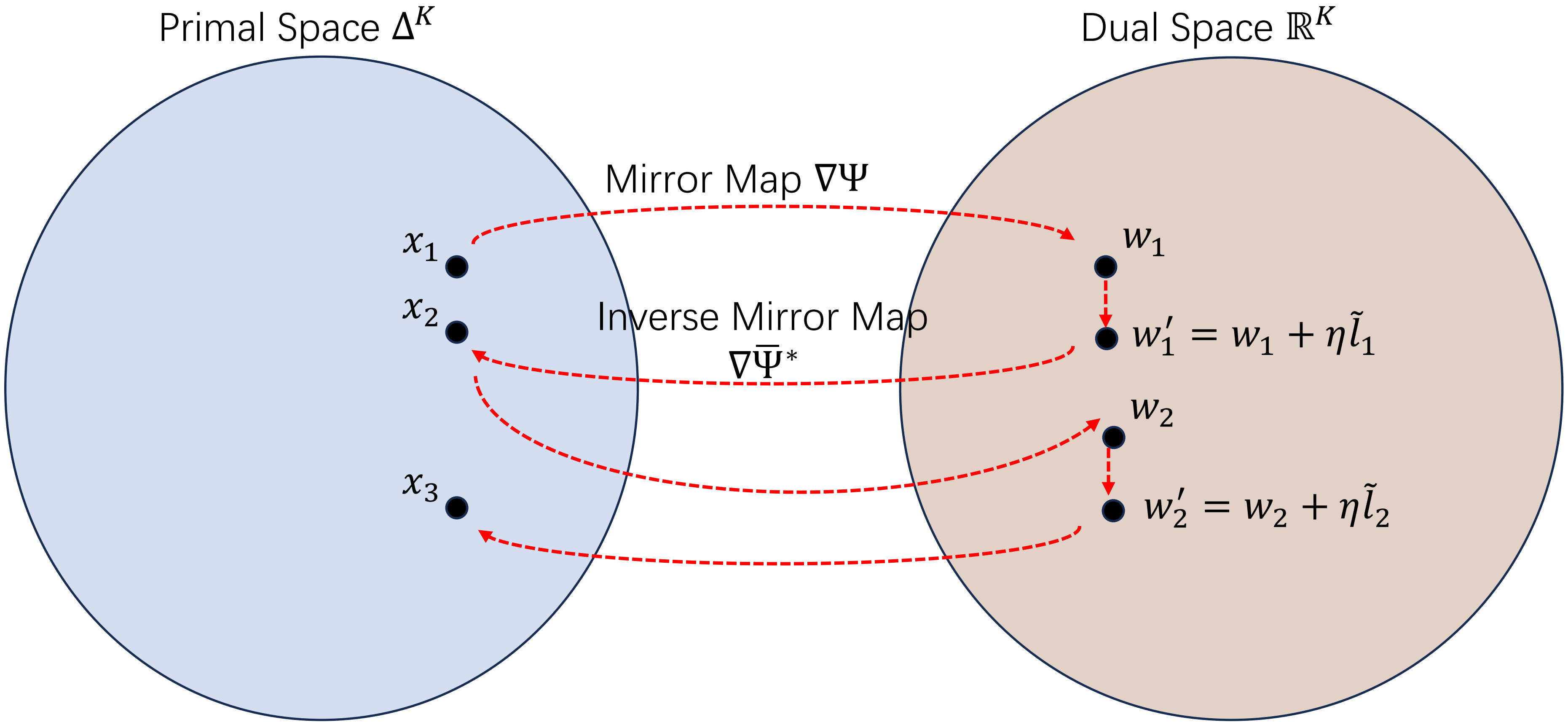
Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
x_1
x_2
w_1
Inverse Mirror Map
$\nabla\bar{\Psi}^*$
$w_1' = w_1 + \eta\tilde{l}_1$
w_2

Vanilla OMD

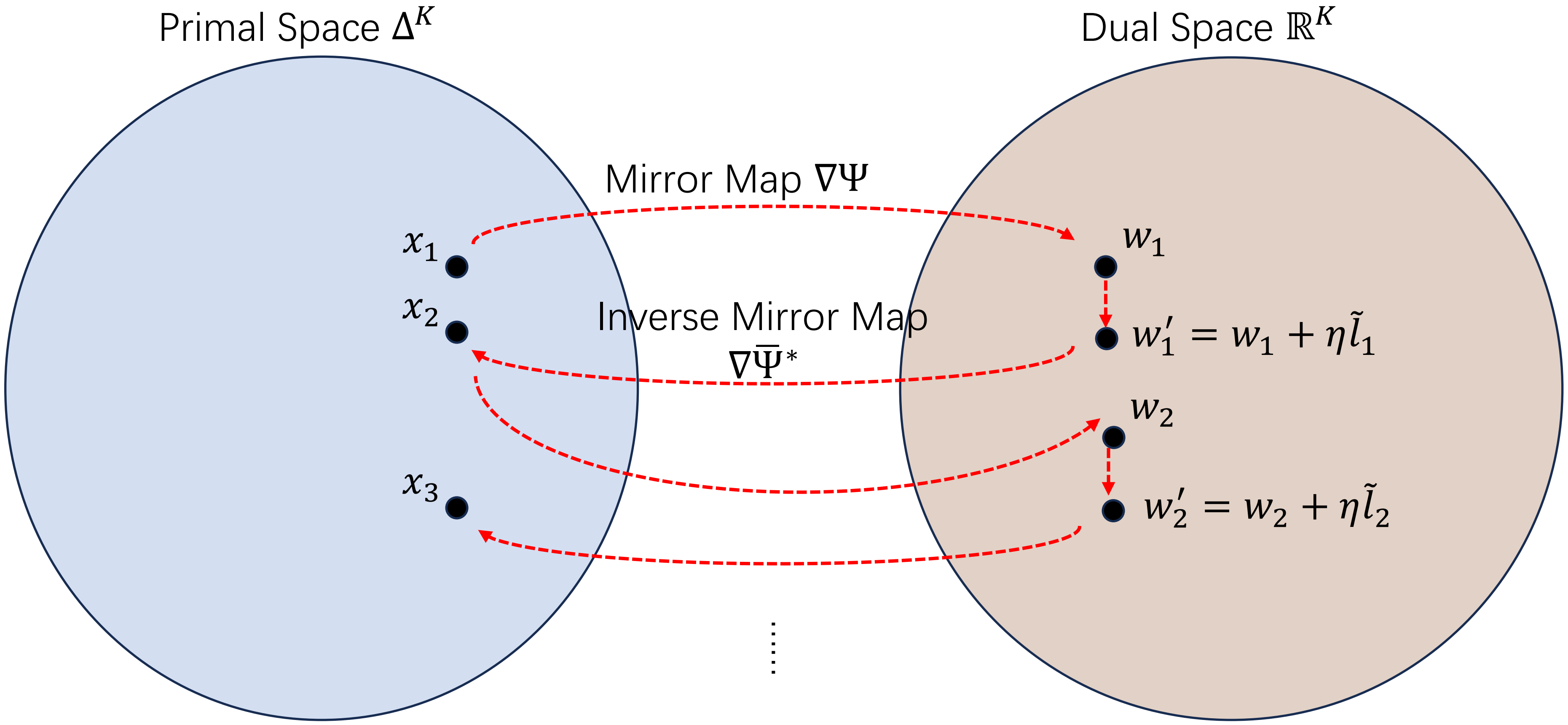


Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
x_1
x_2
w_1
$w_1' = w_1 + \eta\tilde{l}_1$
Inverse Mirror Map
$\nabla\bar{\Psi}^*$
w_2
$w_2' = w_2 + \eta\tilde{l}_2$

Vanilla OMD


Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
Inverse Mirror Map
$\nabla\bar{\Psi}^*$
x_1
x_2
x_3
w_1
$w_1' = w_1 + \eta\tilde{l}_1$
w_2
$w_2' = w_2 + \eta\tilde{l}_2$

Vanilla OMD

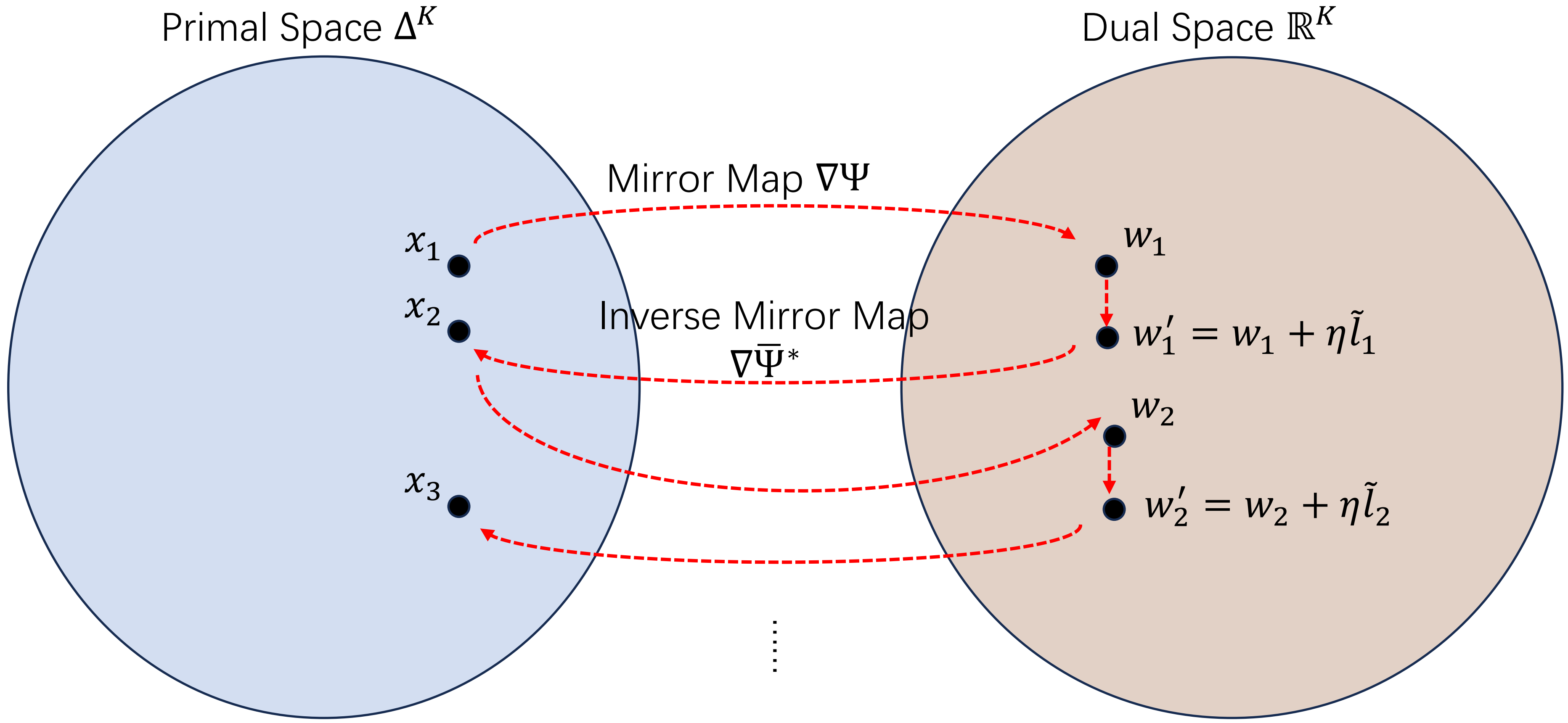


Primal Space Δ^K
Dual Space $\mathbb{R}^K$
Mirror Map $\nabla\Psi$
Inverse Mirror Map
$\nabla\bar{\Psi}^*$
x_1
x_2
x_3
w_1
$w_1' = w_1 + \eta\tilde{l}_1$
w_2
$w_2' = w_2 + \eta\tilde{l}_2$

Vanilla OMD

- Single-step OMD lemma:

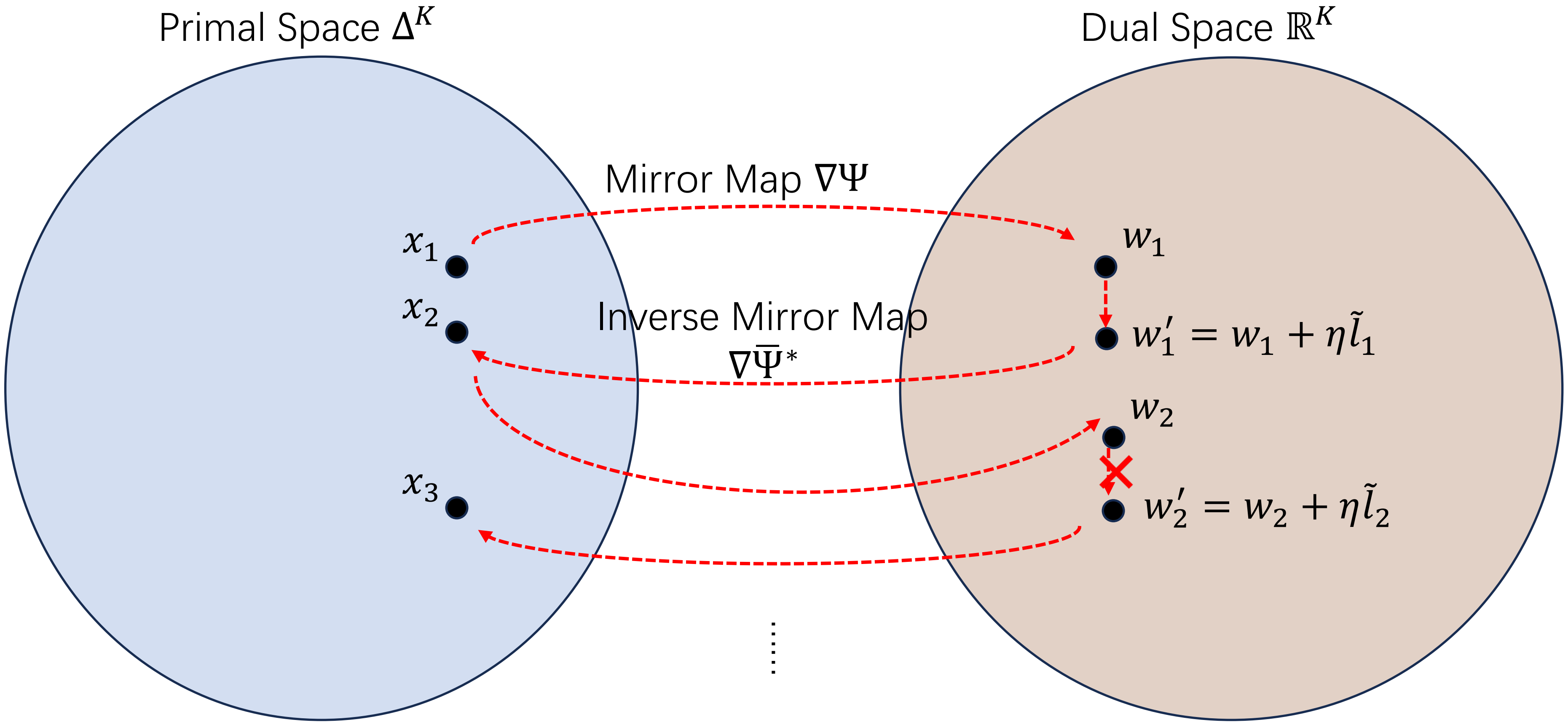
$$\langle x_t - y, \tilde{l}_t \rangle \leq \eta^{-1} D_\Psi(y, x_t) - \eta^{-1} D_\Psi\left(y, \nabla\bar{\Psi}^*({w'}_t)\right) + \eta^{-1} D_{\Psi^*}(w'_t, w_t).$$



Vanilla OMD

- Single-step OMD lemma:

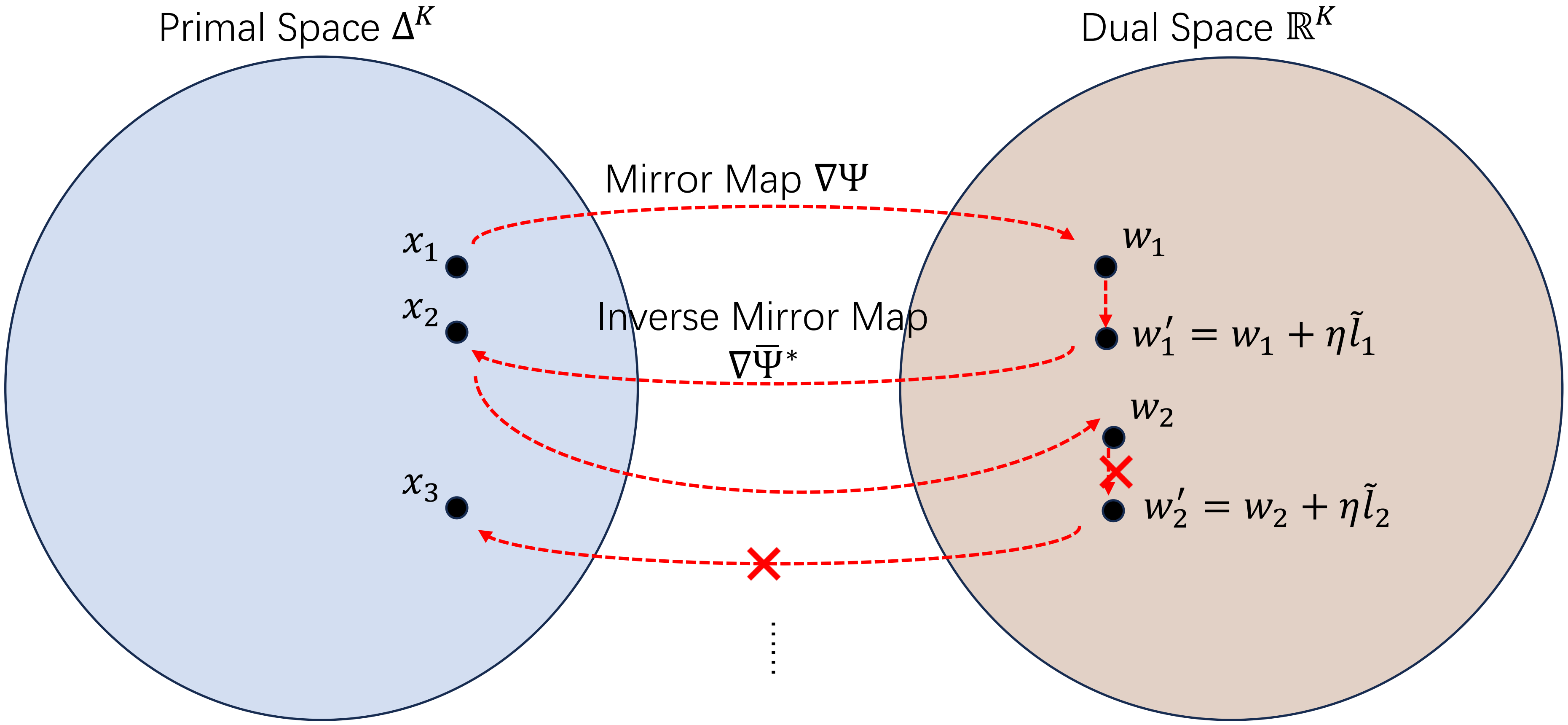
$$\langle x_t - y, \tilde{l}_t \rangle \le \eta^{-1} D_\Psi(y, x_t) - \eta^{-1} D_\Psi\big(y, \nabla\bar{\Psi}^*(w'_t)\big) + \eta^{-1} D_{\Psi^*}(w'_t, w_t).$$



Vanilla OMD

- Single-step OMD lemma:

$$\langle x_t - y, \tilde{l}_t \rangle \le \eta^{-1} D_\Psi(y, x_t) - \eta^{-1} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_t)\right) + \eta^{-1} D_{\Psi^*}(w'_t, w_t).$$



Banker-OMD

Banker-OMD

- A novel framework, generalizing vanilla OMD

Banker-OMD

- A novel framework, generalizing vanilla OMD
- No assumptions on feedback delays and arrival order

Banker-OMD

- A novel framework, generalizing vanilla OMD
- No assumptions on feedback delays and arrival order
  - No words like “feedback of last action”

Banker-OMD

- A novel framework, generalizing vanilla OMD
- No assumptions on feedback delays and arrival order
  - No words like “feedback of last action”
- No assumptions on monotonicity of learning rates

Banker-OMD

- A novel framework, generalizing vanilla OMD
- No assumptions on feedback delays and arrival order
    - No words like “feedback of last action”
- No assumptions on monotonicity of learning rates
- Why Banker?

Banker-OMD

- A novel framework, generalizing vanilla OMD
- No assumptions on feedback delays and arrival order
  - No words like “feedback of last action”
- No assumptions on monotonicity of learning rates
- Why Banker?
  - Fine-grained analysis of potential terms due to OMD steps

High-Level Ideas of Banker-OMD

High-Level Ideas of Banker-OMD

- Calculate w_t' after feedback arrives

High-Level Ideas of Banker-OMD

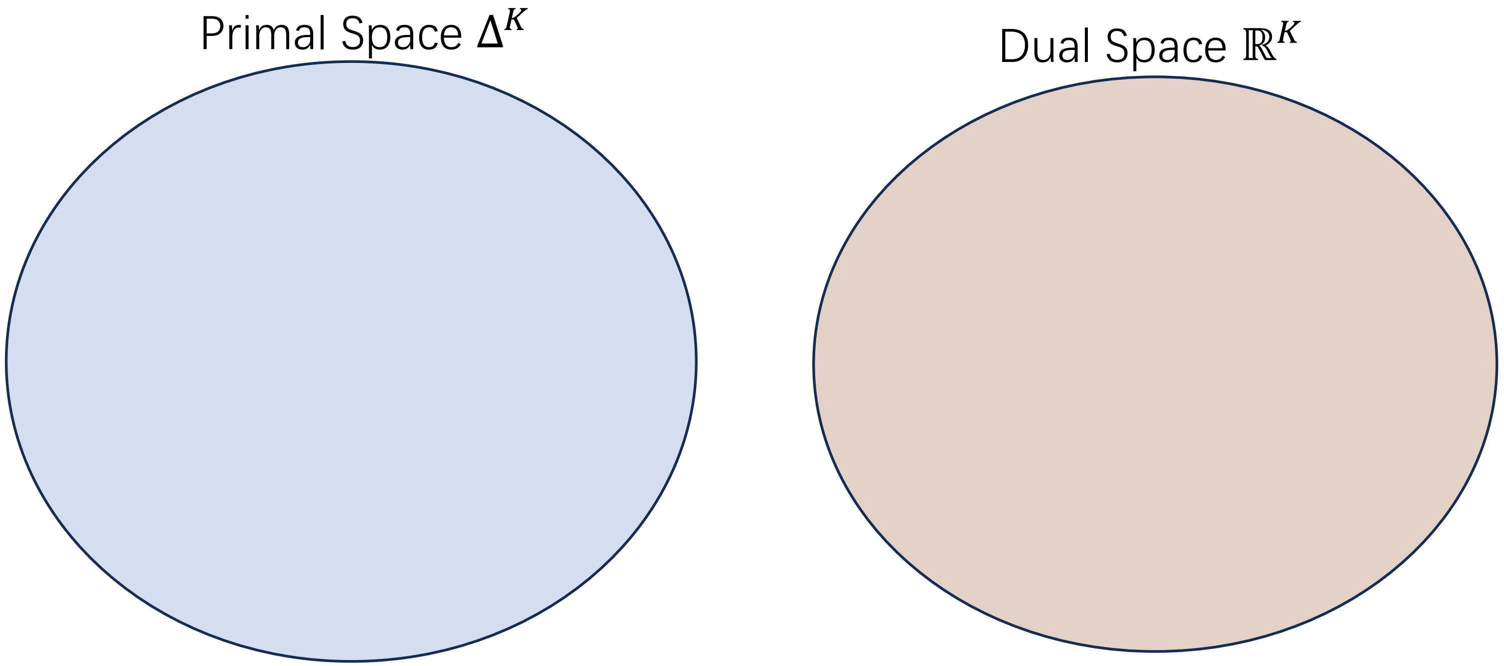
- Calculate w_t' after feedback arrives
- Step-dependent learning rate $\eta_t = \sigma_t^{-1}$

High-Level Ideas of Banker-OMD

- Calculate w_t' after feedback arrives
- Step-dependent learning rate $\eta_t = \sigma_t^{-1}$
  - σ_t "action scale"

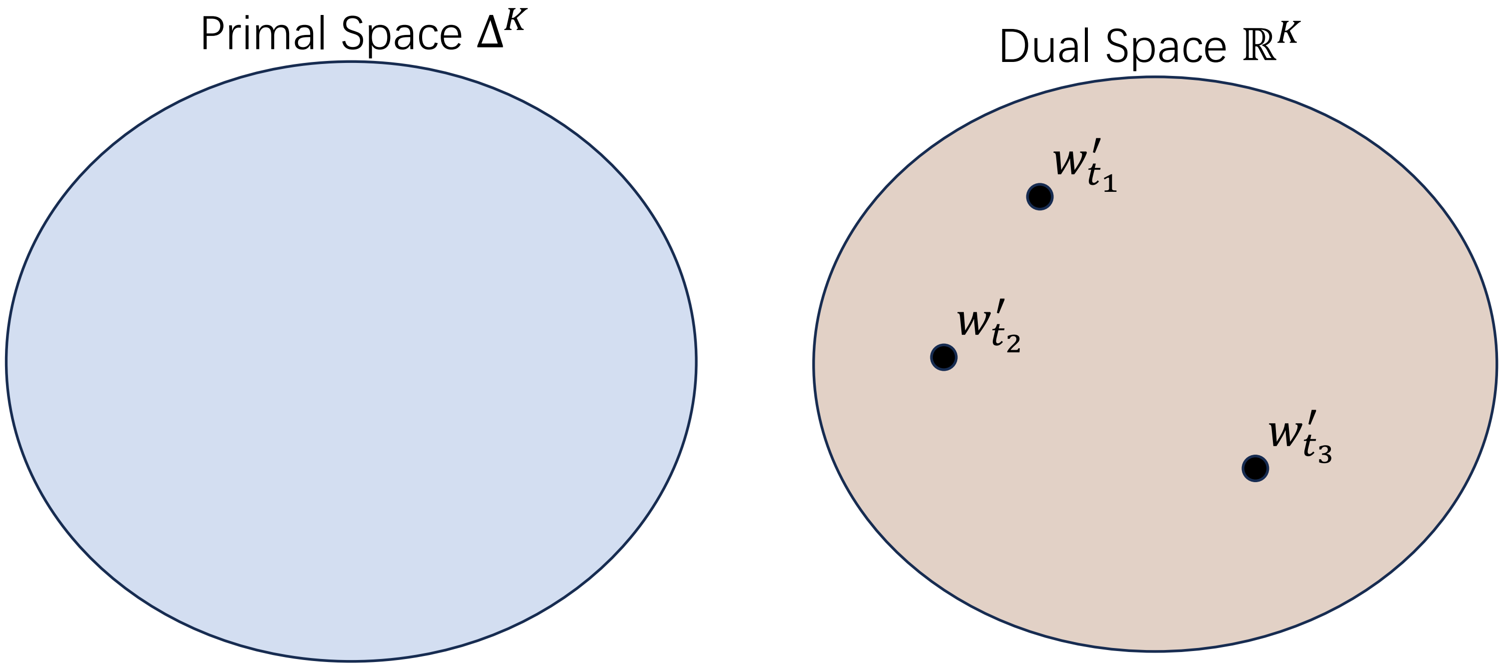
High-Level Ideas of Banker-OMD

- Calculate w_t' after feedback arrives
- Step-dependent learning rate $\eta_t = \sigma_t^{-1}$
  - σ_t “action scale”



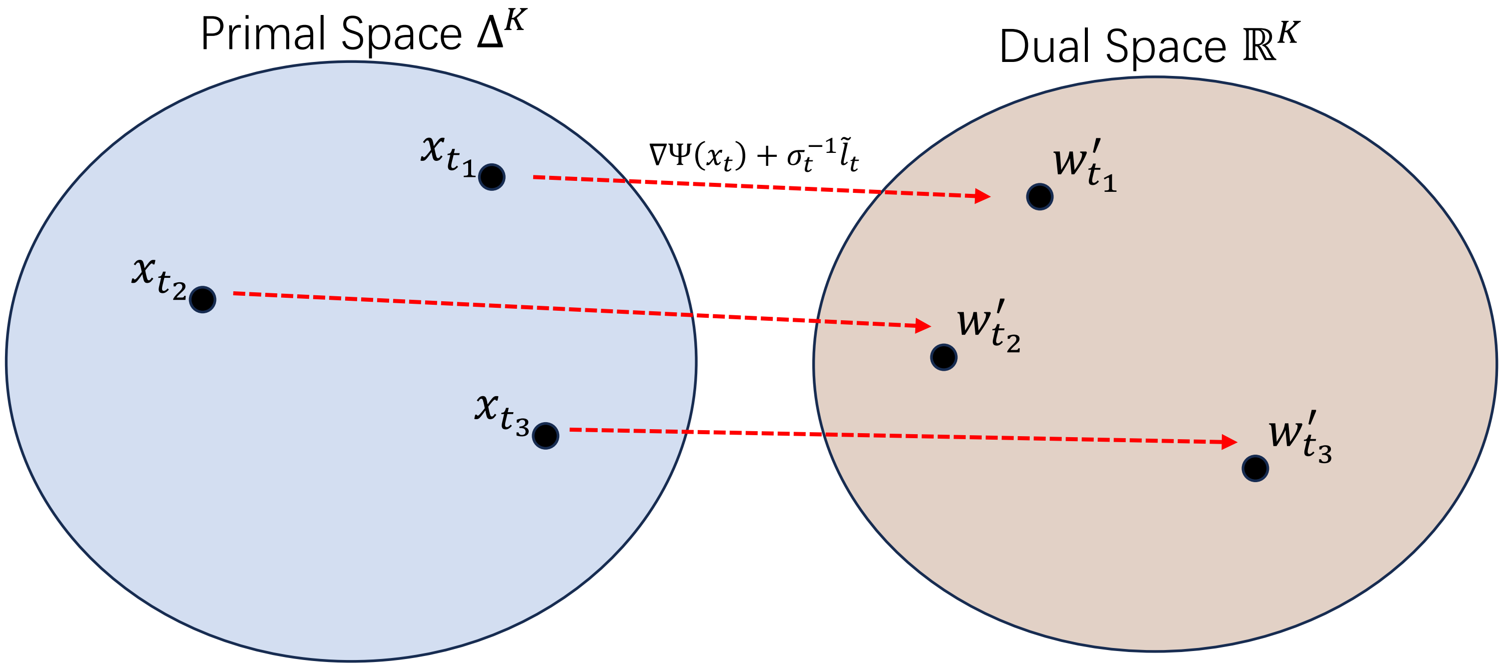
High-Level Ideas of Banker-OMD

- Calculate w'_t after feedback arrives
- Step-dependent learning rate $\eta_t = \sigma_t^{-1}$
  - σ_t "action scale"



High-Level Ideas of Banker-OMD

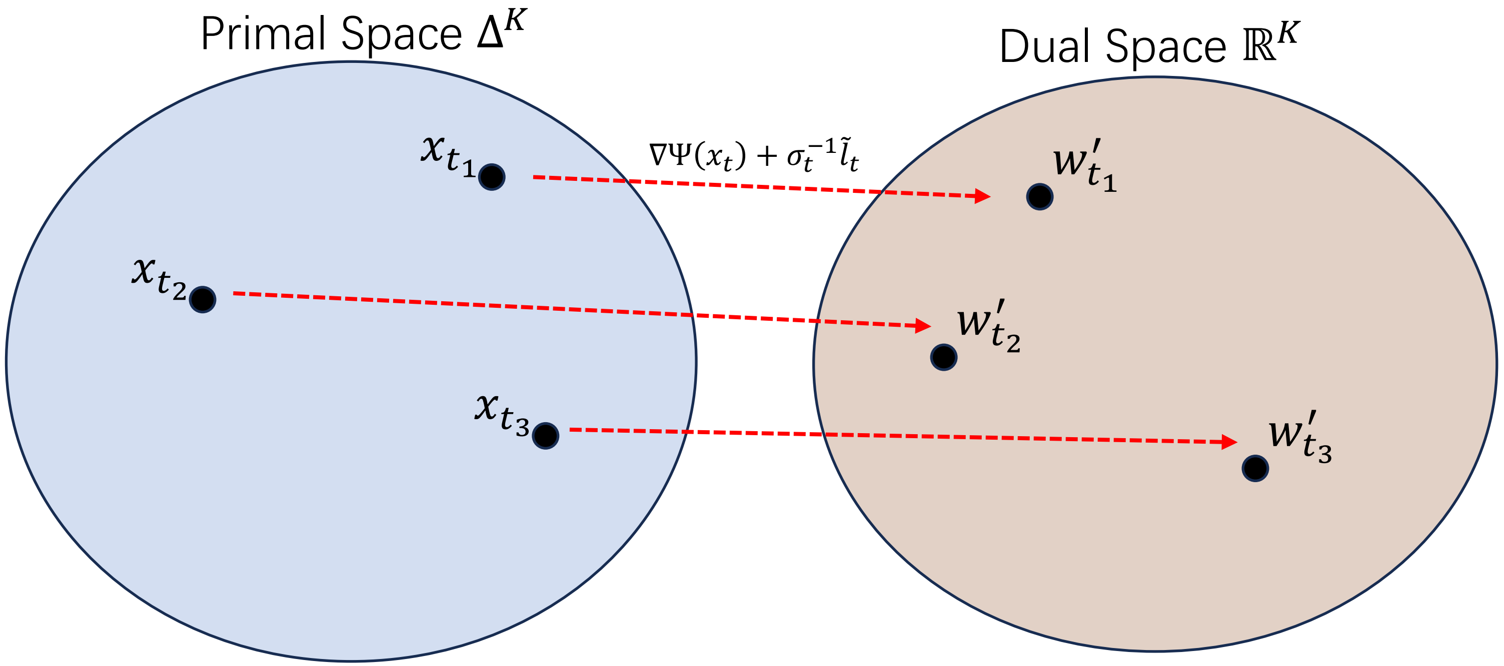
- Calculate w_t' after feedback arrives
- Step-dependent learning rate $\eta_t = \sigma_t^{-1}$
  - σ_t “action scale”



High-Level Ideas of Banker-OMD

- Calculate w'_t after feedback arrives
- Step-dependent learning rate $\eta_t = \sigma_t^{-1}$
  - σ_t "action scale"
- Single-step OMD lemma still holds:

$$\langle x_t - y, \tilde{l}_t \rangle \leq \sigma_t D_\Psi(y, x_t) - \sigma_t D_\Psi\left(y, \nabla\bar{\Psi}^*({w'}_t)\right) + \sigma_t D_{\Psi^*}(w'_t, w_t).$$



High-Level Ideas of Banker-OMD

High-Level Ideas of Banker-OMD

- Core observation:

High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \ldots, t_I$, we have

High-Level Ideas of Banker-OMD

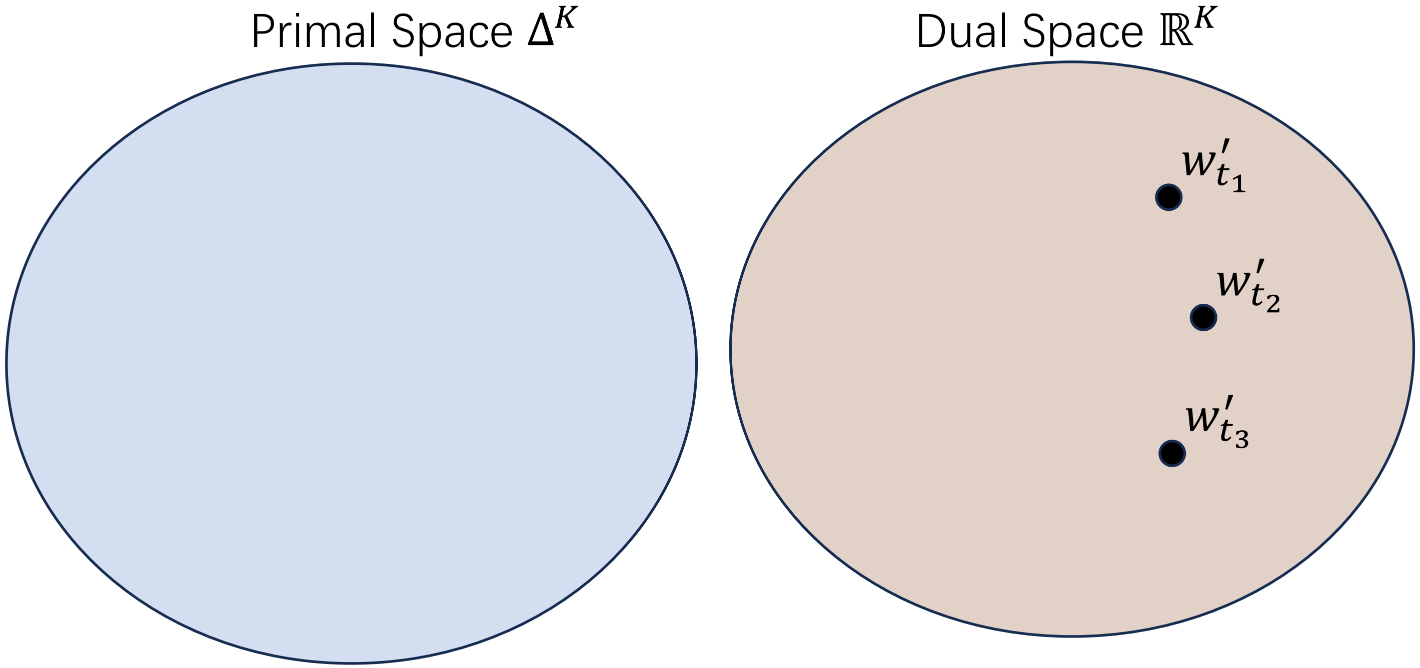
- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \ldots, t_I$, we have

$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$

High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \dots, t_I$, we have

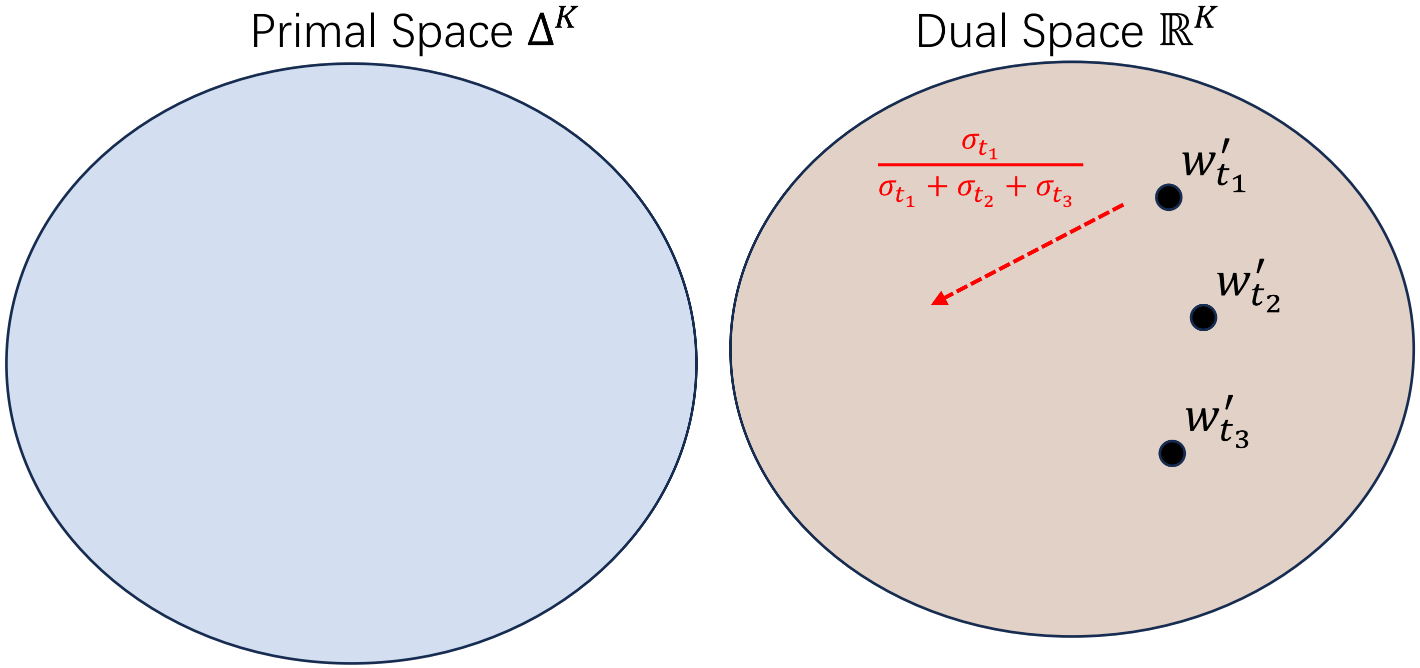
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \ldots, t_I$, we have

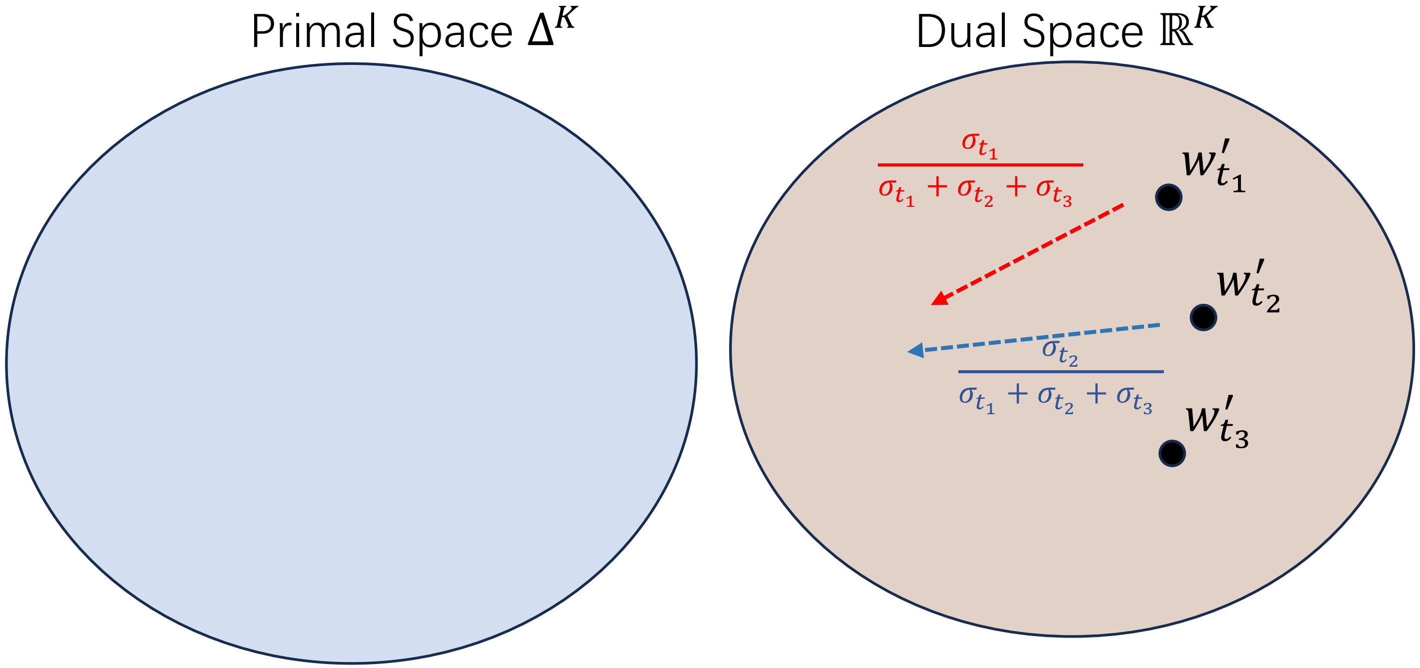
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*\left(w'_{t_i}\right)\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \dots, t_I$, we have

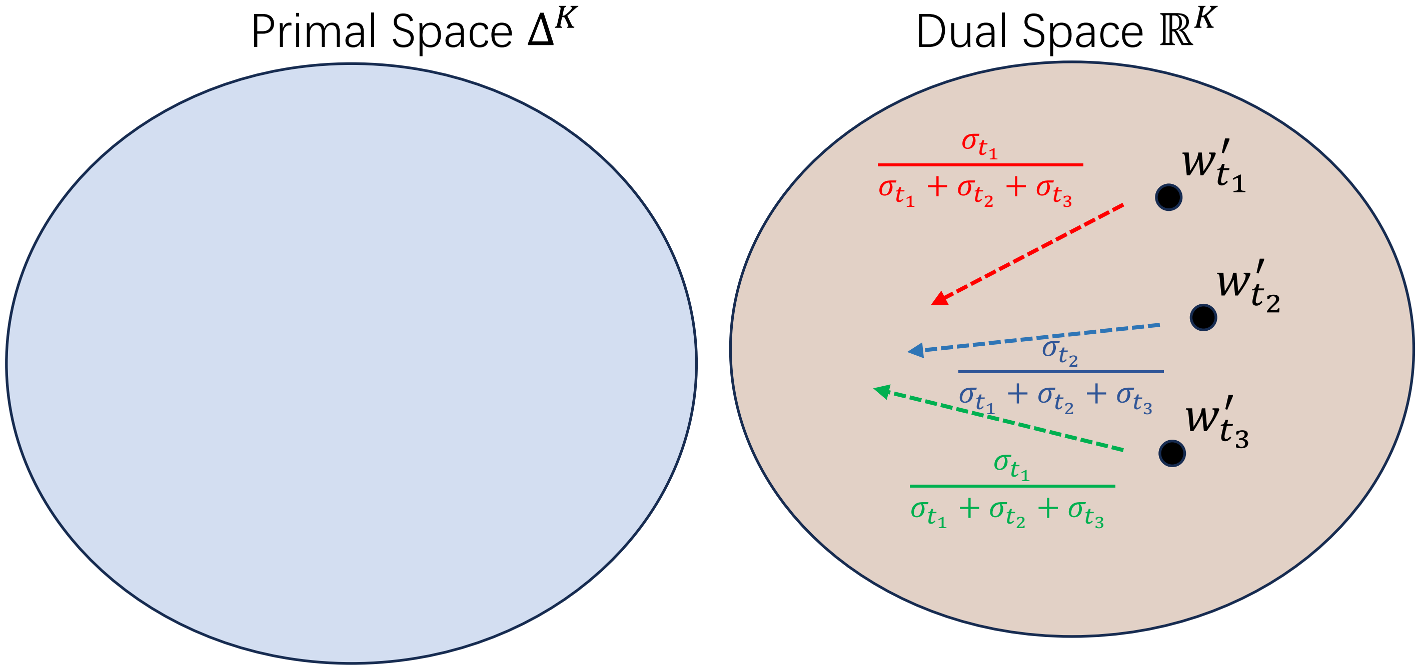
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \ldots, t_I$, we have

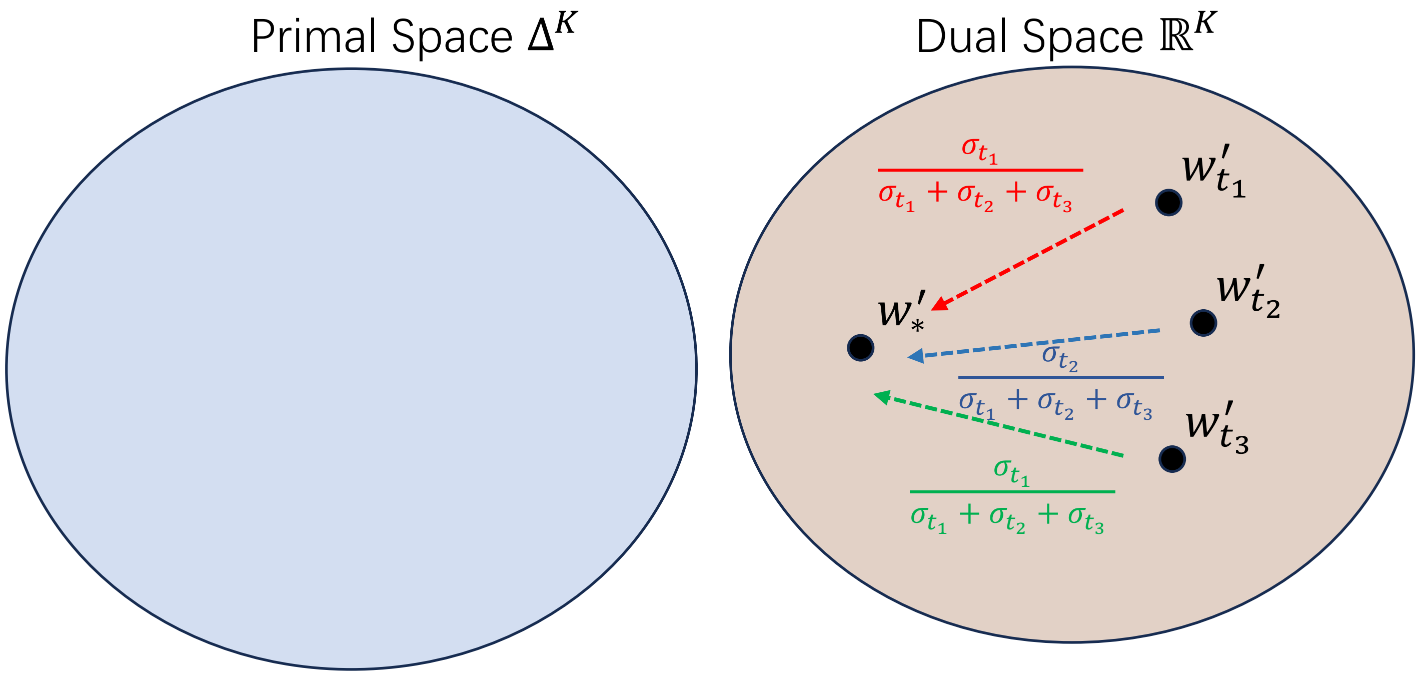
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \dots, t_I$, we have

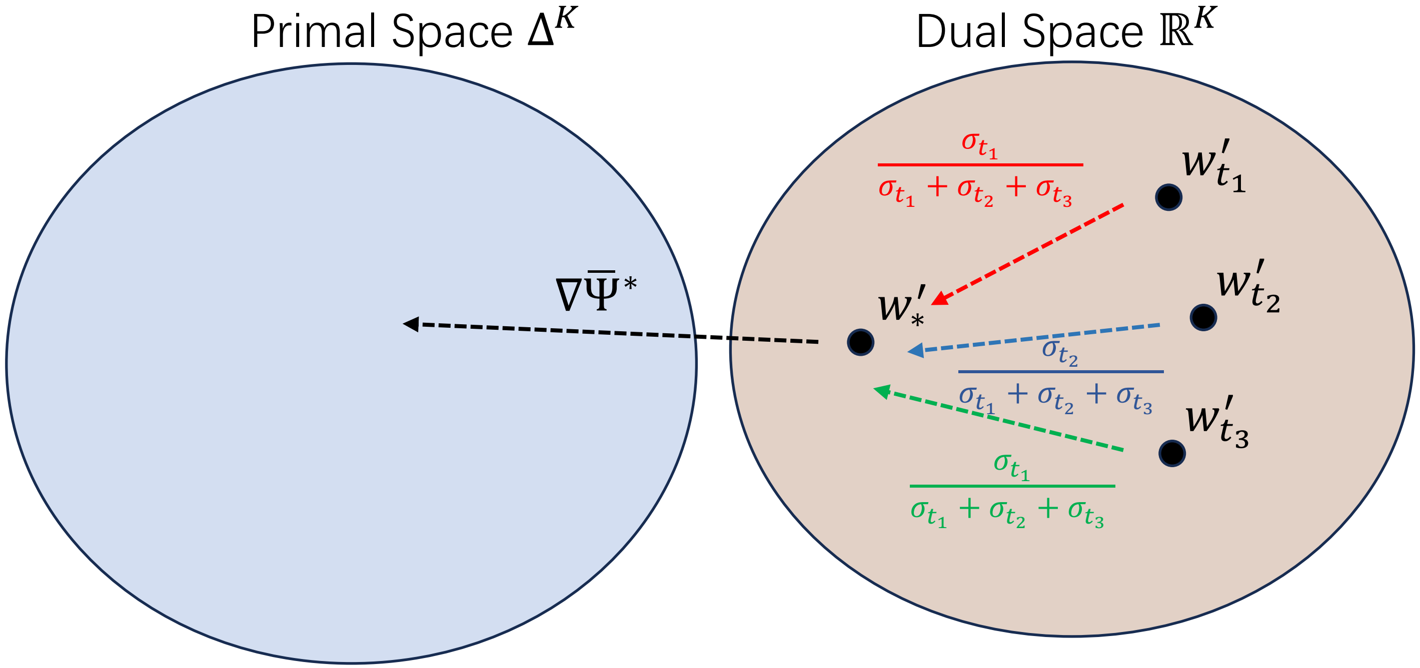
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \dots, t_I$, we have

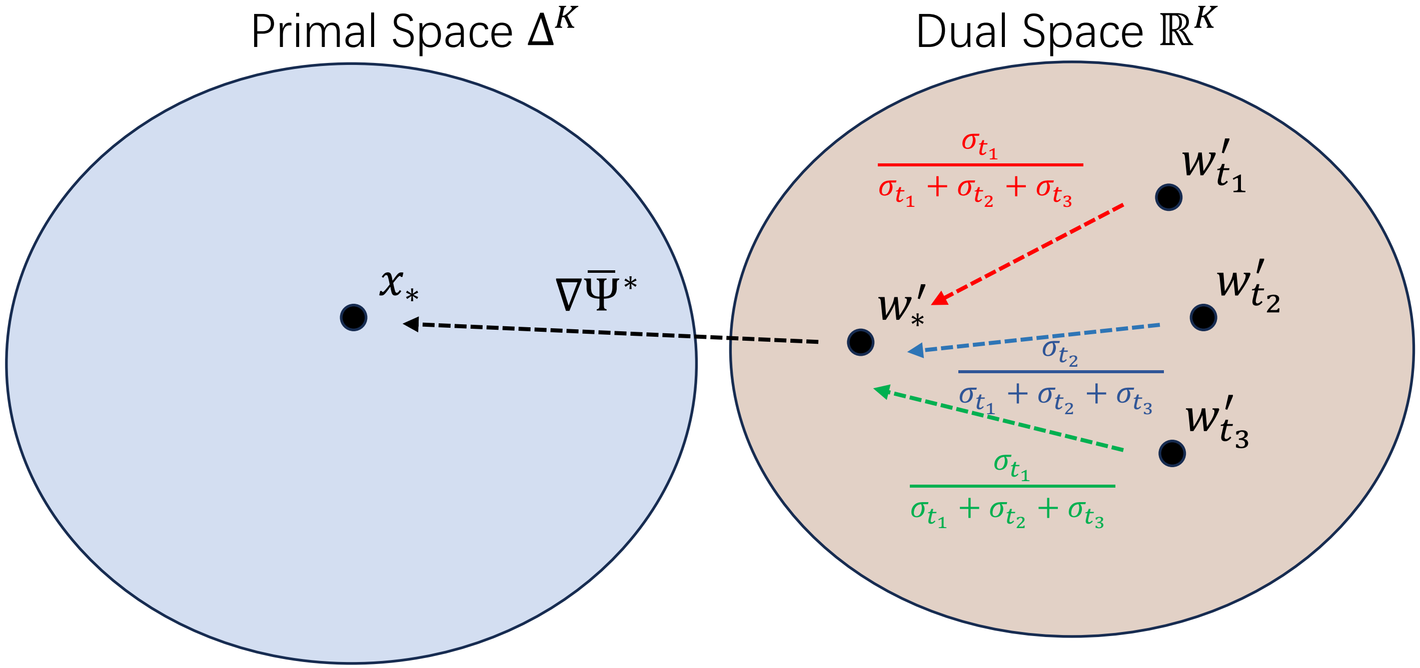
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \dots, t_I$, we have

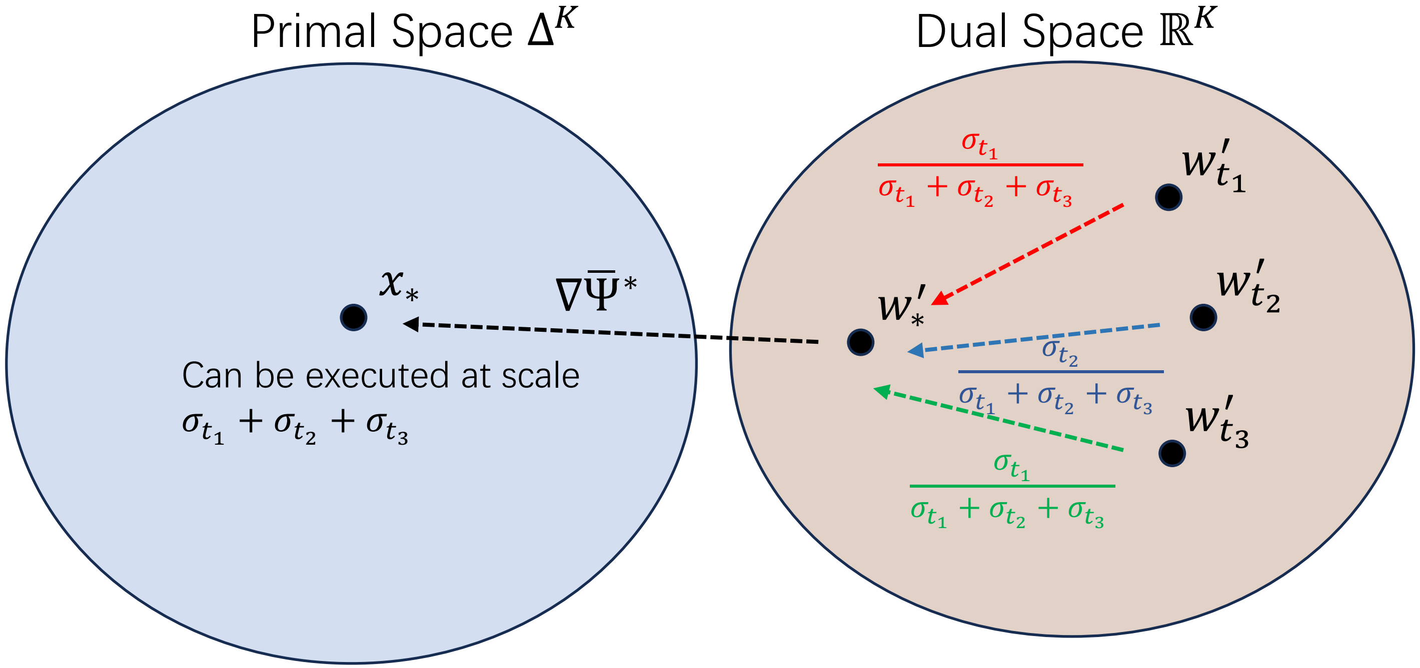
$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \ge \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$



High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \ldots, t_I$, we have

$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$

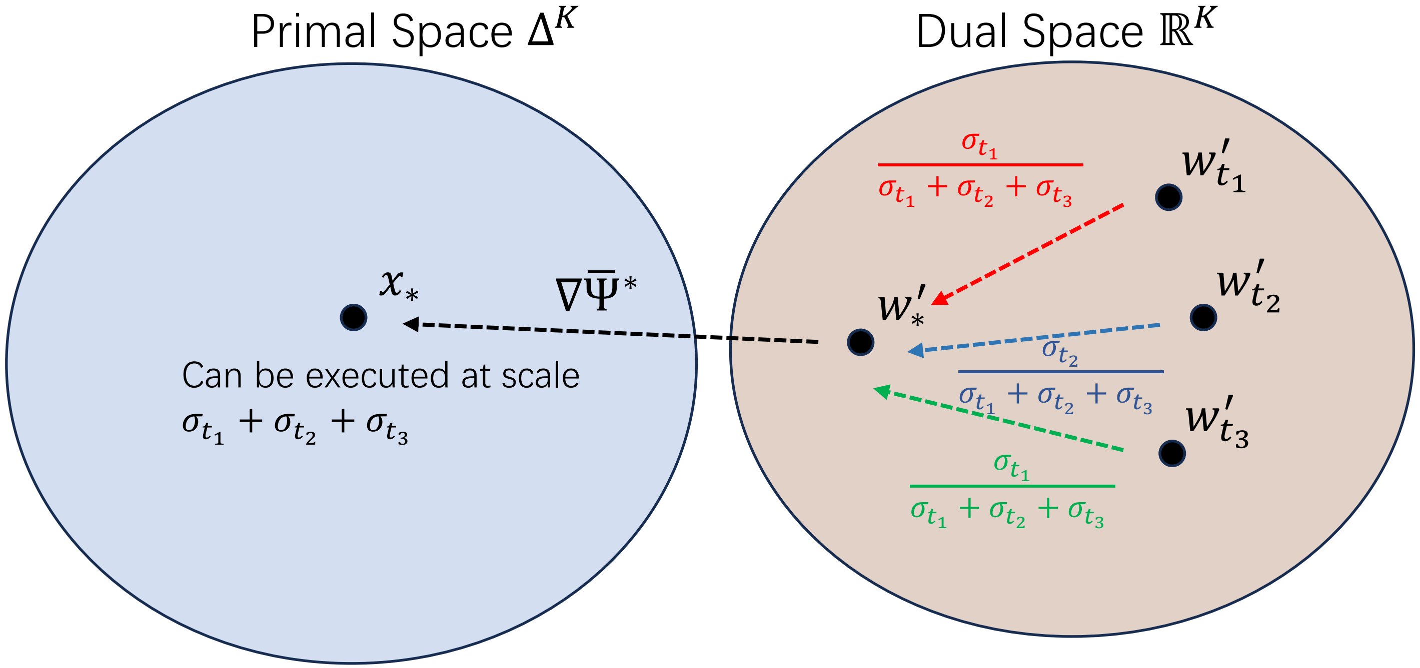


High-Level Ideas of Banker-OMD

- Core observation:
  - Convex combination on dual space keeps balance of bookkeeping: $\forall t_1, t_2, \ldots, t_I$, we have

$$\sum_i \sigma_{t_i} D_\Psi\left(y, \nabla\bar{\Psi}^*(w'_{t_i})\right) \geq \sigma_\Sigma D_\Psi(y, x_*), \qquad \text{where } \sigma_\Sigma = \sum_i \sigma_{t_i}, x_* = \nabla\bar{\Psi}^*\left(\sum_i \frac{\sigma_{t_i}}{\sigma_\Sigma} w'_{t_i}\right).$$

  - We are allowed to execute x^* at scale σ_Σ "**free of charge**"!



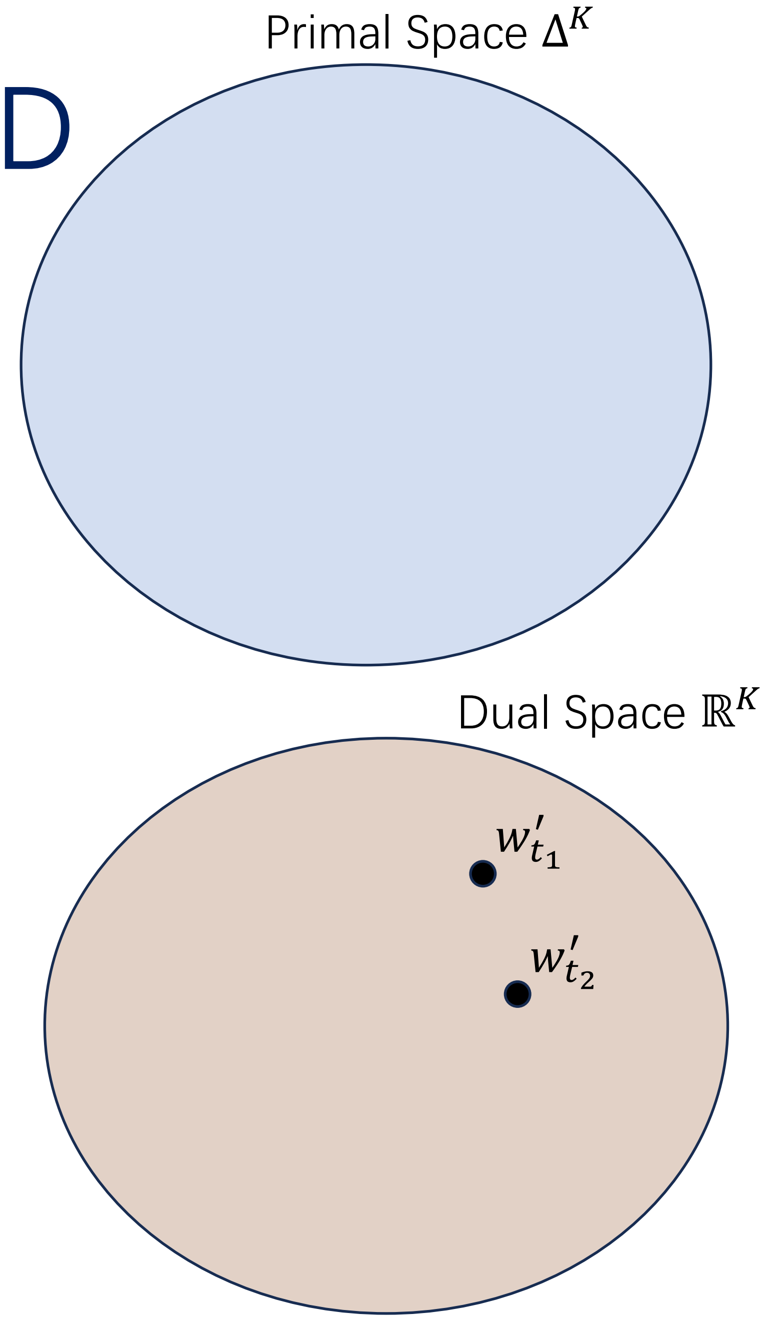
High-Level Ideas of Banker-OMD

High-Level Ideas of Banker-OMD

- Overdrafting:

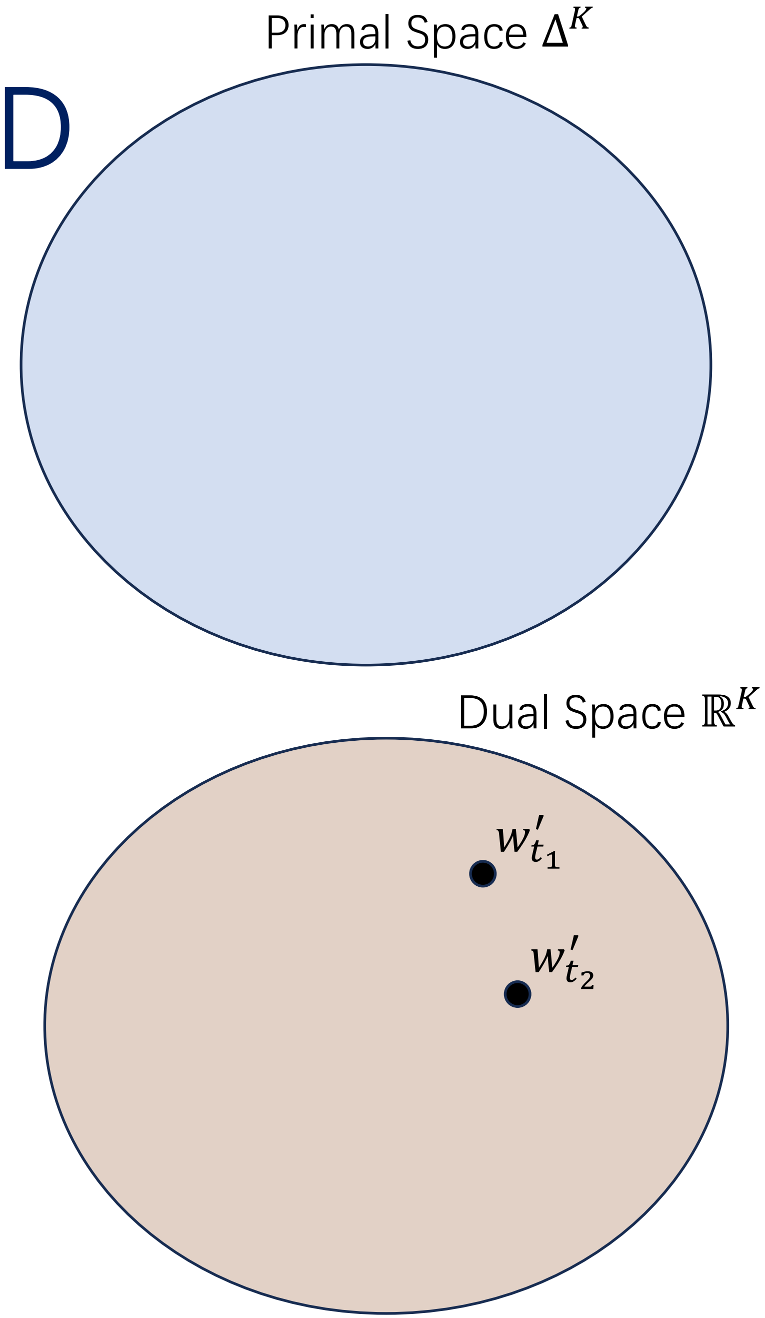
High-Level Ideas of Banker-OMD

- Overdrafting:



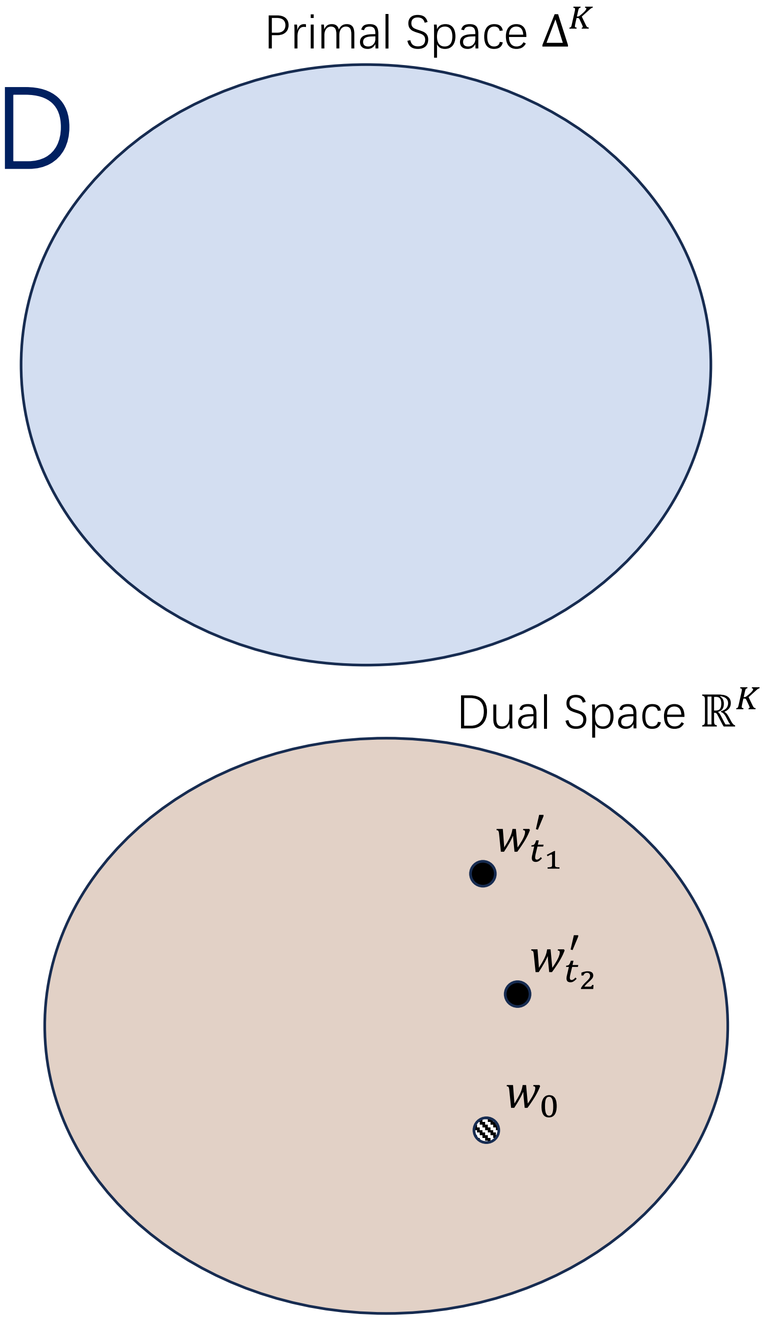
High-Level Ideas of Banker-OMD

- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_{\sum} = \sigma_{t_1} + \sigma_{t_2}$?



High-Level Ideas of Banker-OMD

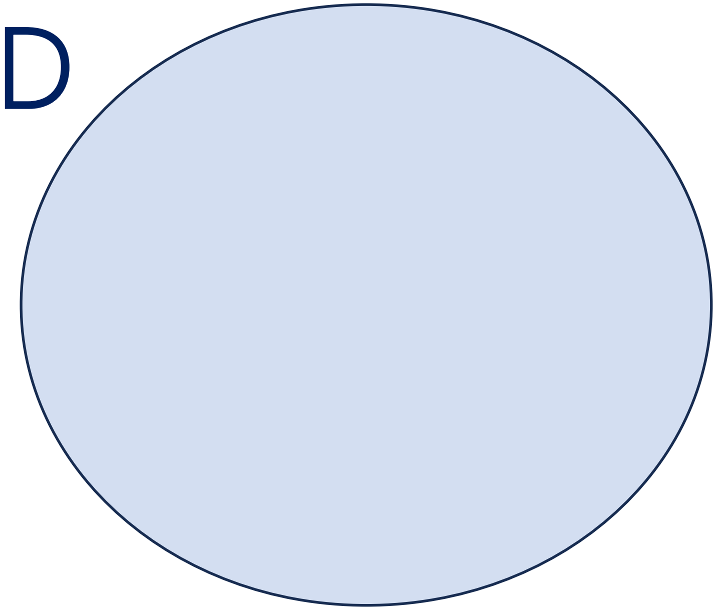
- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_{\sum} = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)



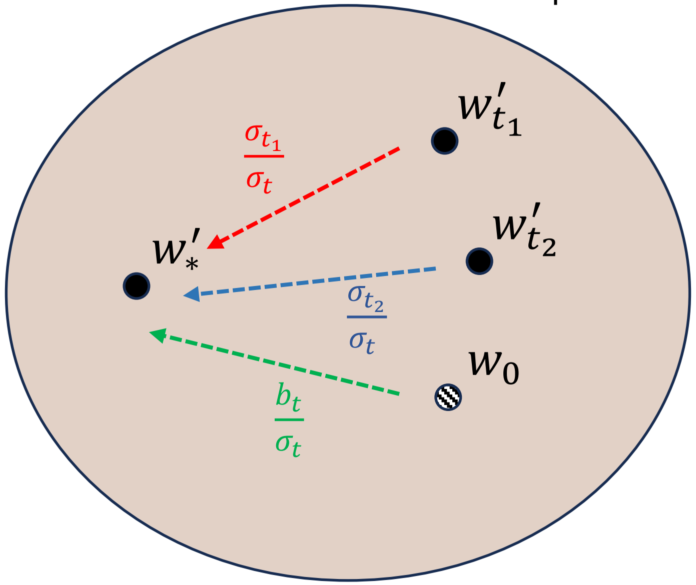
High-Level Ideas of Banker-OMD

- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$

Primal Space Δ^K

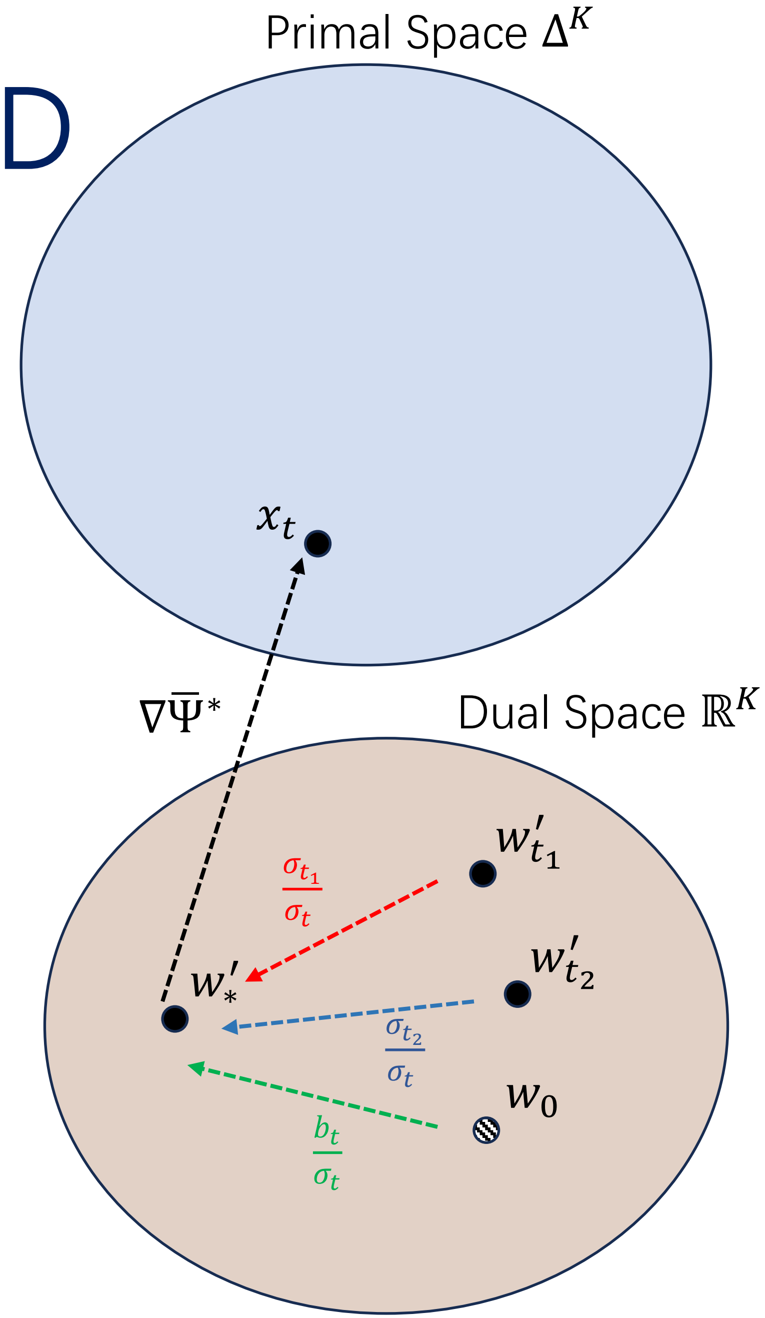


Dual Space $\mathbb{R}^K$



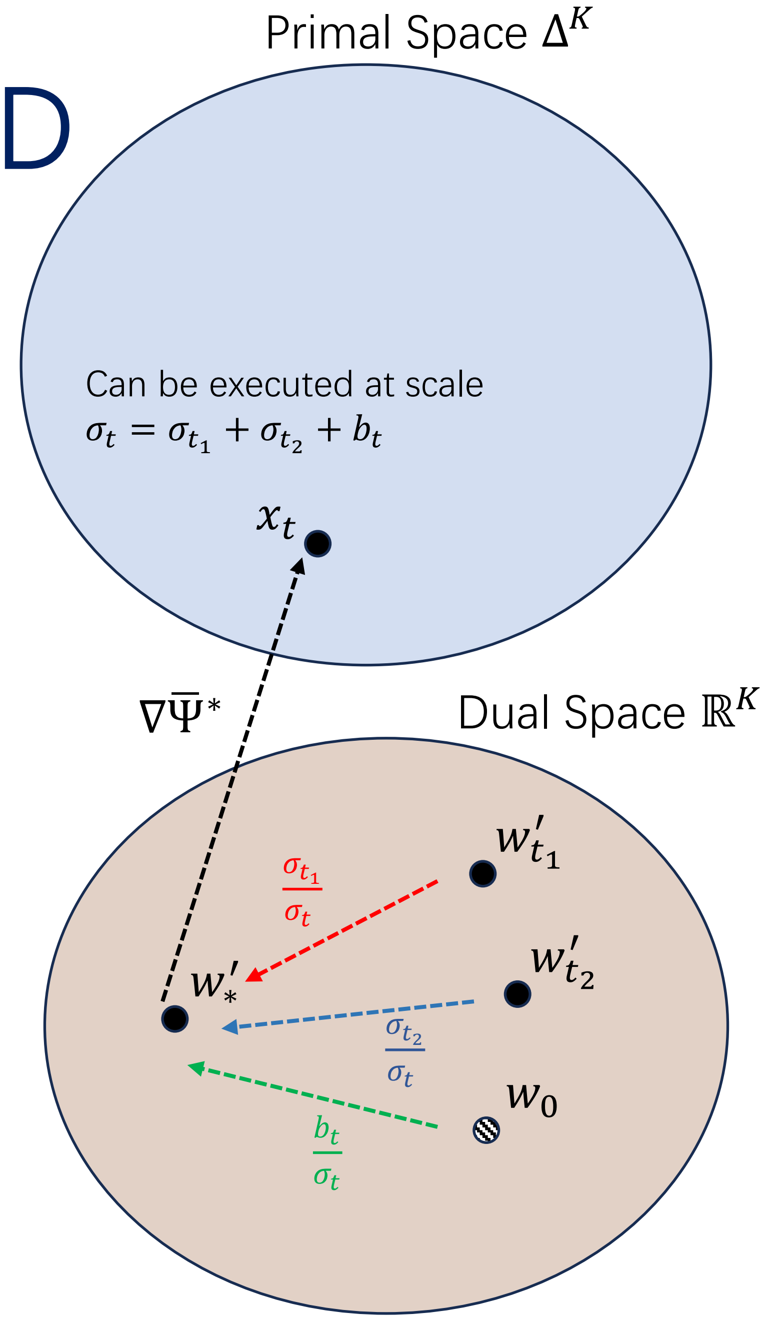
High-Level Ideas of Banker-OMD

- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$



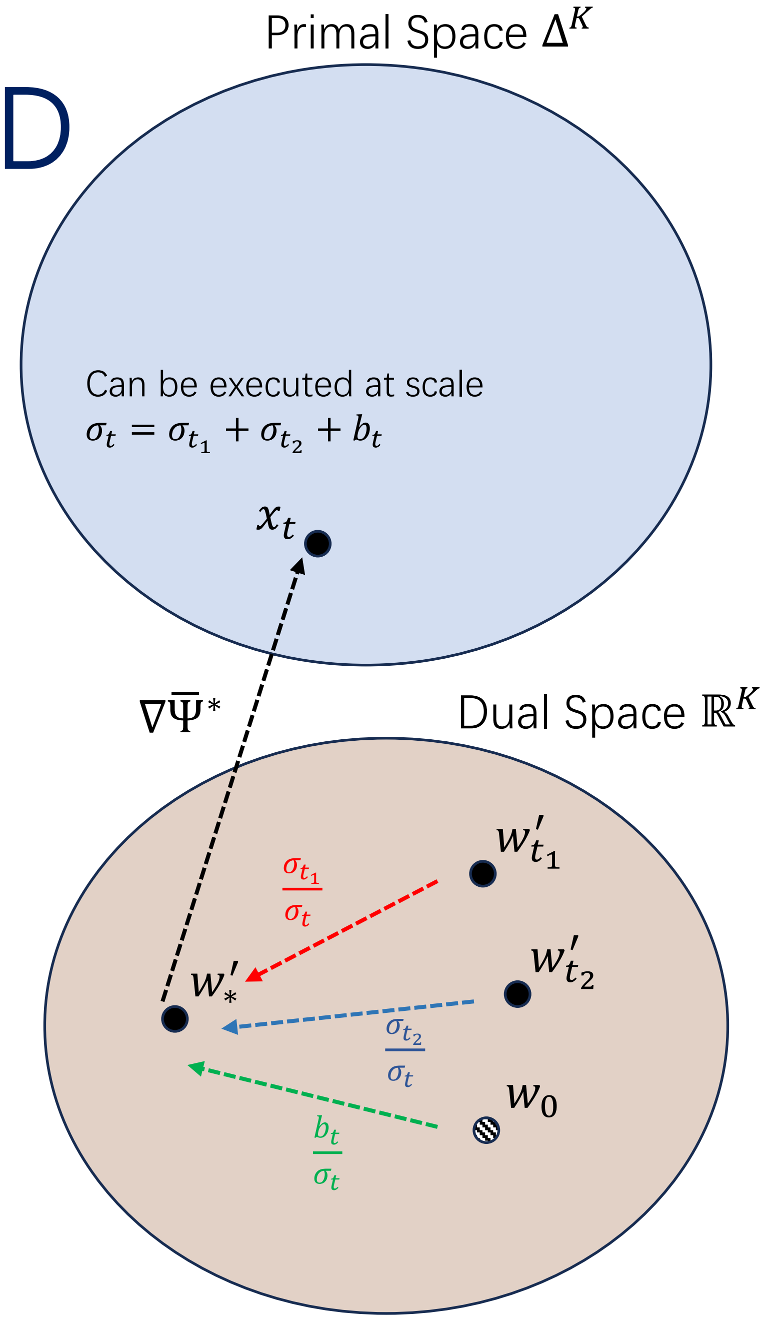
High-Level Ideas of Banker-OMD

- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$



High-Level Ideas of Banker-OMD

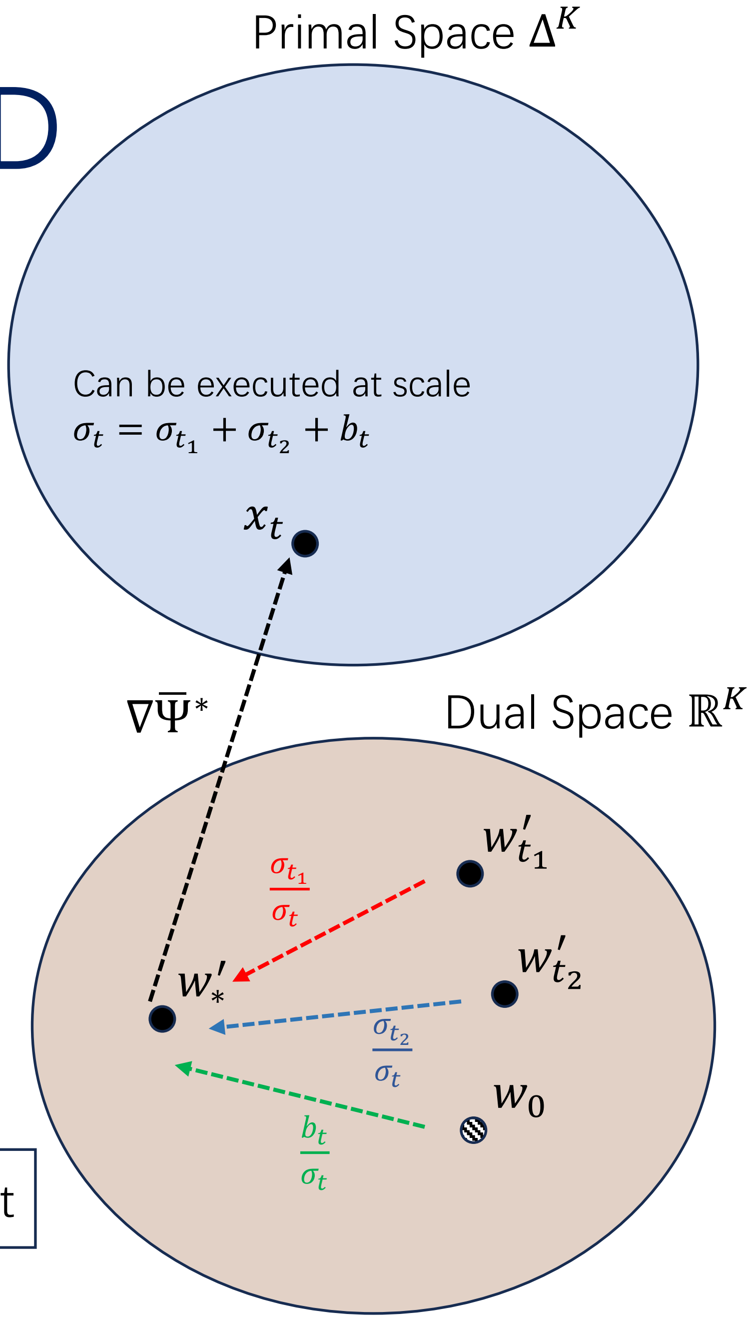
- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a “default investment” $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required “investment” on x_0: $b_t = \sigma_t - \sigma_\Sigma$
  - “Imaginary” $b_t D_\Psi(y, x_0) - b_t D_\Psi\left(y, \nabla\bar{\Psi}^*(w_0)\right)$ terms



High-Level Ideas of Banker-OMD

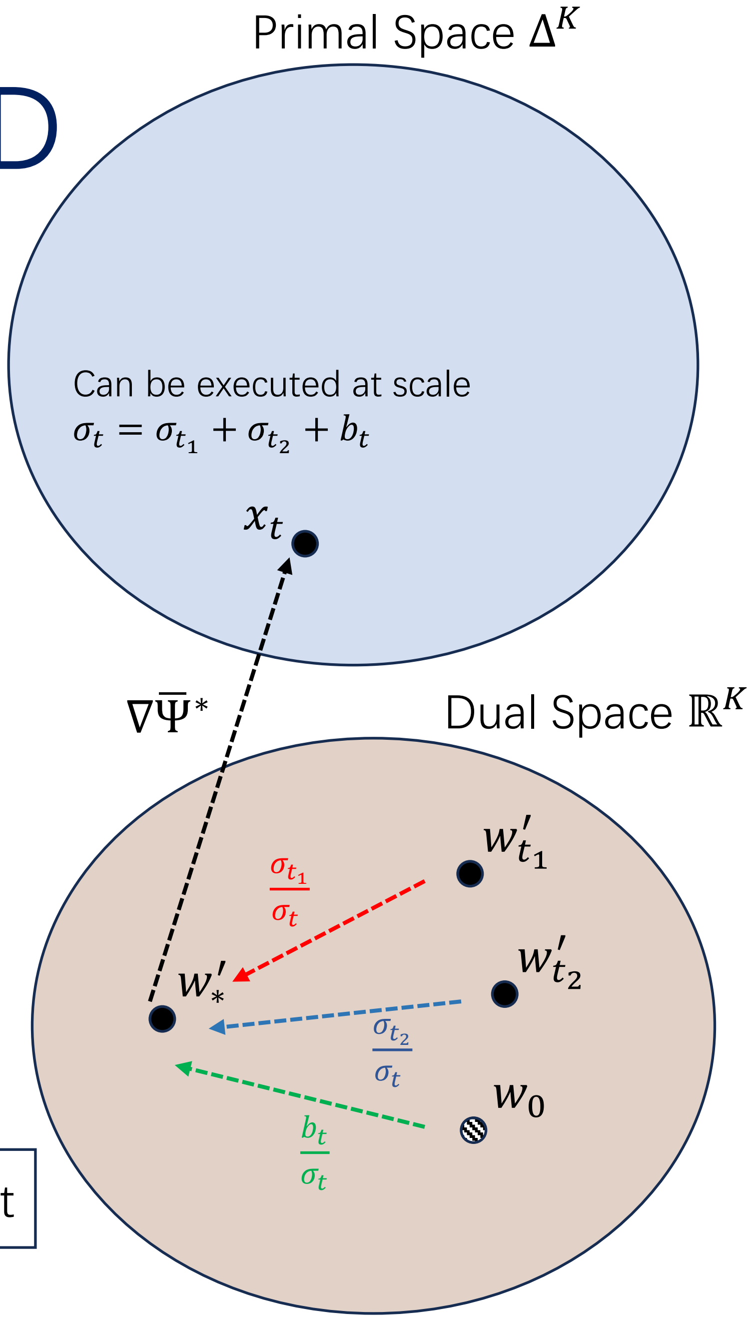
- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$
  - "Imaginary" $b_t D_\Psi(y, x_0) - b_t D_\Psi\left(y, \nabla\bar{\Psi}^*(w_0)\right)$ terms

$b_t D_\Psi(y, x_0)$ extra cost



High-Level Ideas of Banker-OMD

- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \ldots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$
  - "Imaginary" $b_t D_\Psi(y, x_0) - b_t D_\Psi\left(y, \nabla\bar{\Psi}^*(w_0)\right)$ terms
- Banker-OMD:



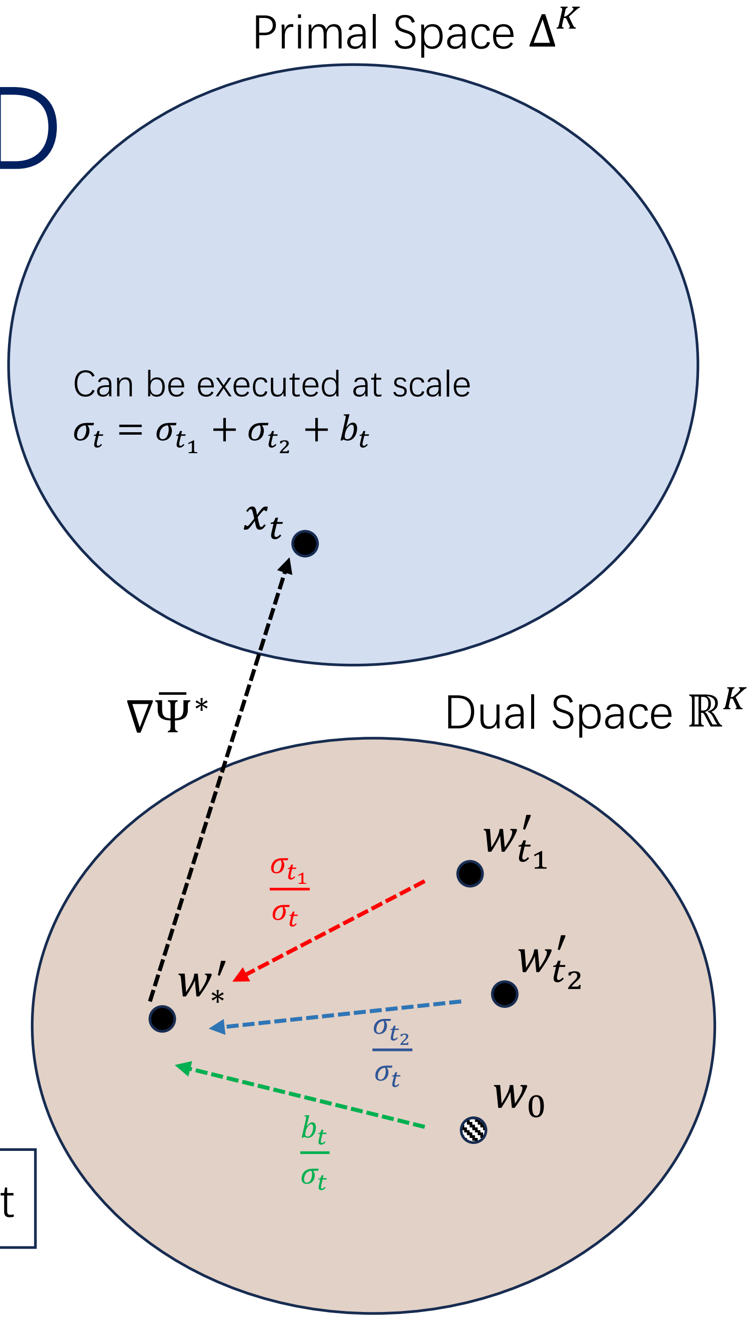
$b_t D_\Psi(y, x_0)$ extra cost

High-Level Ideas of Banker-OMD

- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$
  - "Imaginary" $b_t D_\Psi(y, x_0) - b_t D_\Psi\left(y, \nabla\bar{\Psi}^*(w_0)\right)$ terms
- Banker-OMD:
  - Consistent rule for regret bookkeeping, ensuring

$$\text{Regret}_T \le \sum_t b_t \cdot D_\Psi(y, x_0) + \sum_t \sigma_t D_{\Psi^*}(w_t', w_t)\,!$$

$b_t D_\Psi(y, x_0)$ extra cost



High-Level Ideas of Banker-OMD

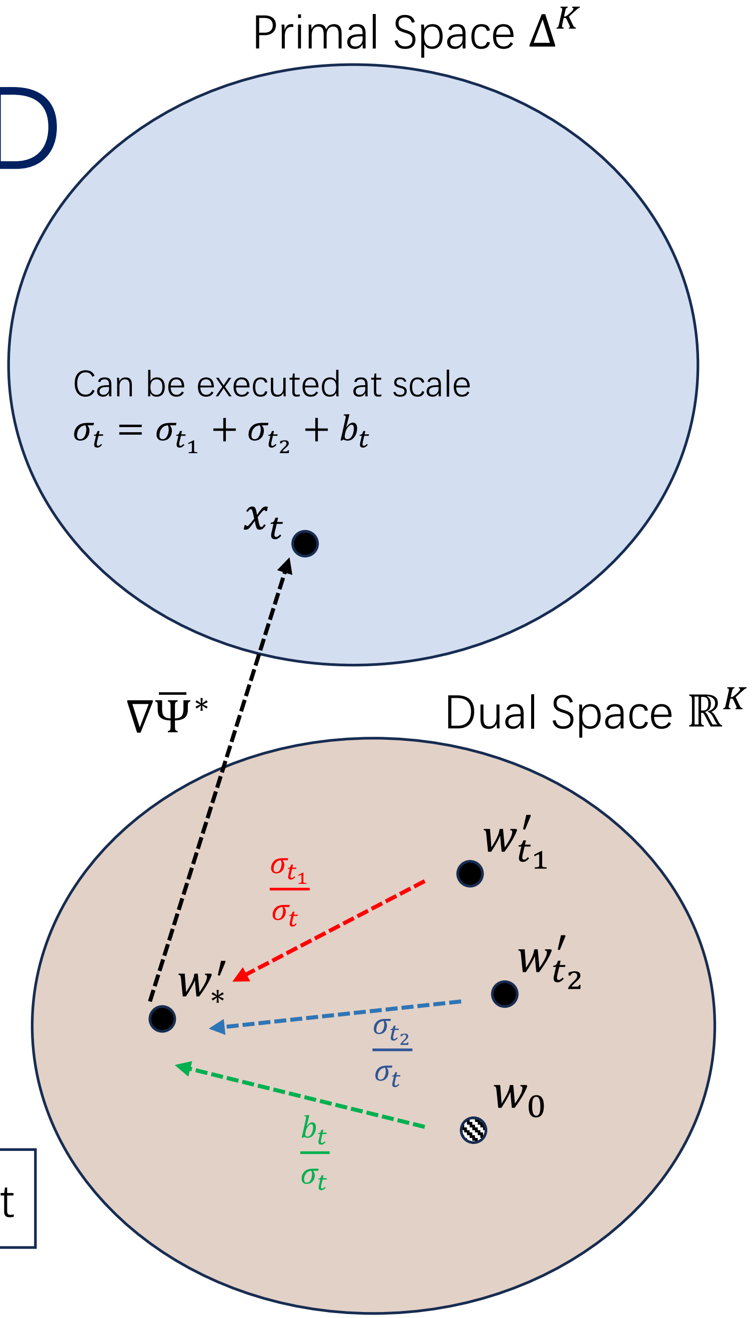
- Overdrafting:
  - Want if we want larger scale $\sigma_t > \sigma_\Sigma = \sigma_{t_1} + \sigma_{t_2}$?
  - Apply a "default investment" $x_0 = \left(\frac{1}{K}, \dots, \frac{1}{K}\right)$ (with mirror image w_0)
  - Required "investment" on x_0: $b_t = \sigma_t - \sigma_\Sigma$
  - "Imaginary" $b_t D_\Psi(y, x_0) - b_t D_\Psi\left(y, \nabla\bar{\Psi}^*(w_0)\right)$ terms
- Banker-OMD:
  - Consistent rule for regret bookkeeping, ensuring

$$\mathrm{Regret}_T \leq \sum_t b_t \cdot D_\Psi(y, x_0) + \sum_t \sigma_t D_{\Psi^*}(w_t', w_t)\,!$$

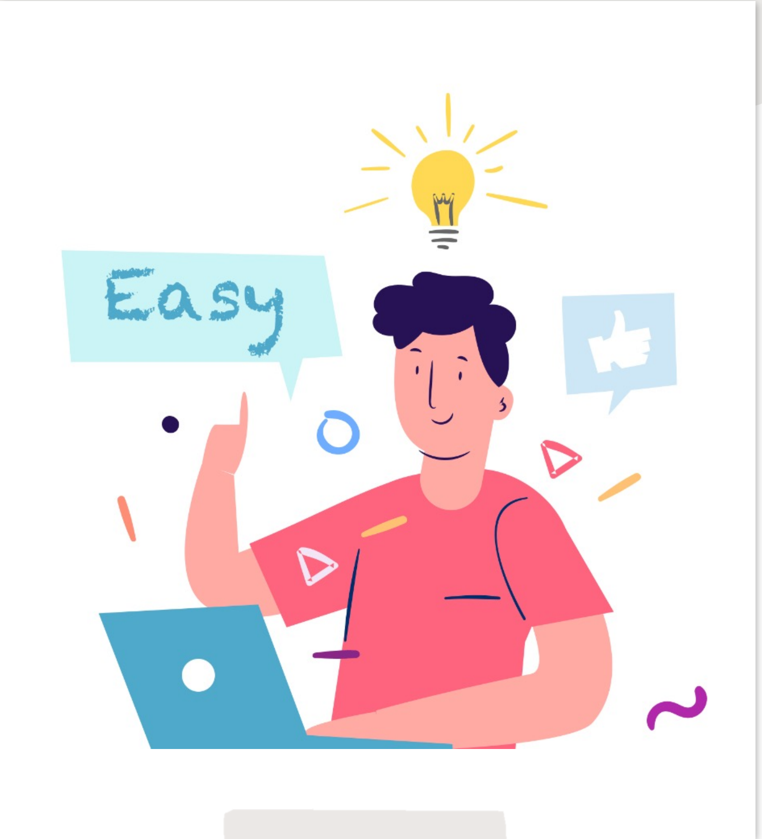
  - And... provides general scale rule to deal with delays!

$$\tilde{\mathcal{O}}\left(\sqrt{D+T}\right) - \text{style bounds made easy!}$$

$b_t D_\Psi(y, x_0)$ extra cost



Main Theorem of Banker-OMD

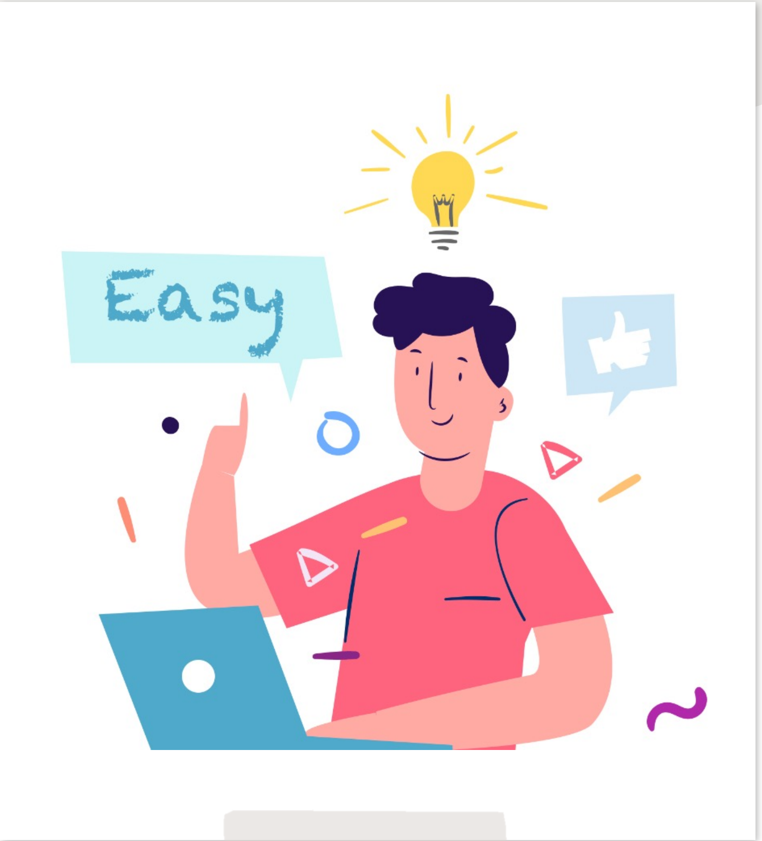


- Given a practical algorithm based on vanilla OMD with $\mathcal{O}(C\sqrt{T})$ regret for non-delayed adversarial bandit problem, there is a Banker-OMD based version using the same regularizer, guaranteeing

$$\mathcal{O}\left(C\sqrt{T} + C'\sqrt{D\log D}\right)$$

regret in the delayed-feedback setting.

Main Theorem of Banker-OMD



- Given a practical algorithm based on vanilla OMD with $\mathcal{O}(C\sqrt{T})$ regret for non-delayed adversarial bandit problem, there is a Banker-OMD based version using the same regularizer, guaranteeing

$$\mathcal{O}\left(C\sqrt{T} + C'\sqrt{D\log D}\right)$$

regret in the delayed-feedback setting.

- **Non-delayed** Algorithm ≈ **OMD** + Regularizer + Step-sizes
- **Delay-robust** Algorithm ≈ **Banker-OMD**+ Same regularizer + Modified step-sizes

New Results of Banker-OMD

New Results of Banker-OMD

- BOLO **(Abernethy et al., 2008)** ensures regret

$$O\left(n^{1.5}\sqrt{T\log T}\right)$$

for n-dim adversarial linear bandits.

New Results of Banker-OMD

- BOLO **(Abernethy et al., 2008)** ensures regret

$$O\left(n^{1.5}\sqrt{T\log T}\right)$$

for n-dim adversarial linear bandits.

- Banker-BOLO **(Ours)** ensures regret

$$\mathcal{O}\left(n^{1.5}\sqrt{\log T}\left(\sqrt{T}+\sqrt{D\log D}\right)+n^2\sqrt{D}\log T\right)$$

for n-dim delayed adversarial linear bandits.

New Results of Banker-OMD

- BOLO **(Abernethy et al., 2008)** ensures regret
  $$O\left(n^{1.5}\sqrt{T\log T}\right)$$
  for n-dim adversarial linear bandits.
- Banker-BOLO **(Ours)** ensures regret
  $$\mathcal{O}\left(n^{1.5}\sqrt{\log T}\left(\sqrt{T}+\sqrt{D\log D}\right)+n^2\sqrt{D}\log T\right)$$
  for n-dim delayed adversarial linear bandits.
- State-of-the-art regret bound for non-delayed scale-free MABs **(Ours):**
  $$\mathcal{O}\left(\sqrt{KT}L\log T+L\log L\right).$$

New Results of Banker-OMD

- BOLO **(Abernethy et al., 2008)** ensures regret
$$O\left(n^{1.5}\sqrt{T\log T}\right)$$
for n-dim adversarial linear bandits.
- Banker-BOLO **(Ours)** ensures regret
$$\mathcal{O}\left(n^{1.5}\sqrt{\log T}\left(\sqrt{T}+\sqrt{D\log D}\right)+n^2\sqrt{D}\log T\right)$$
for n-dim delayed adversarial linear bandits.
- State-of-the-art regret bound
for non-delayed scale-free MABs **(Ours):**
$$\mathcal{O}\left(\sqrt{KT}L\log T+L\log L\right).$$
- Banker version regret bound
for delayed scale-free MABs **(Ours):**
$$\tilde{O}\left(\sqrt{K(D+T)L}\right).$$

The End

- Thank for listening!

References

Putta S R, Agrawal S. Scale-Free Adversarial Multi Armed Bandits[C]//International Conference on Algorithmic Learning Theory. PMLR, 2022: 910-930.

Bistritz I, Zhou Z, Chen X, et al. Online exp3 learning in adversarial bandits with delayed feedback[J]. Advances in neural information processing systems, 2019, 32.

Abernethy J D, Hazan E, Rakhlin A. An efficient algorithm for bandit linear optimization[C]//21st Annual Conference on Learning Theory. 2008.

Zimmert J, Seldin Y. An optimal algorithm for adversarial bandits with arbitrary delays[C]//International Conference on Artificial Intelligence and Statistics. PMLR, 2020: 3285-3294.

Thune T S, Cesa-Bianchi N, Seldin Y. Nonstochastic multiarmed bandits with unrestricted delays[J]. Advances in Neural Information Processing Systems, 2019, 32.