Generalized Majorization-Minimization





Sobhan Naderi
Google Research
(Presenter)

Kun He
Facebook Reality Labs

Reza Aghajani
UCSD

Stan Sclaroff
Boston University

Pedro Felzenszwalb
Brown University

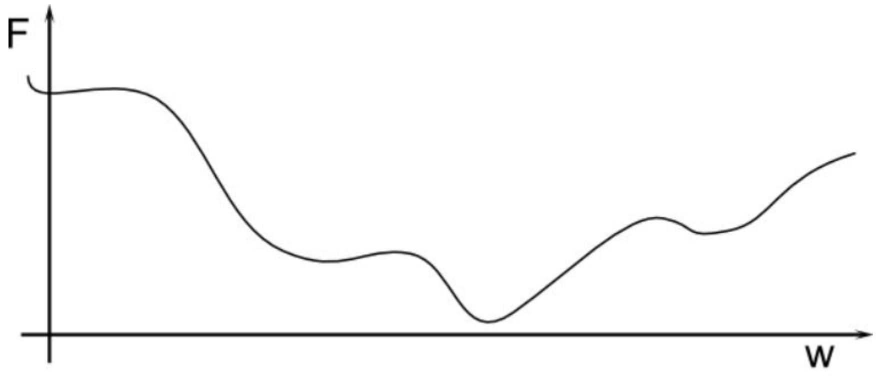
ICML 2019 Long Beach, CA, USA

Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)

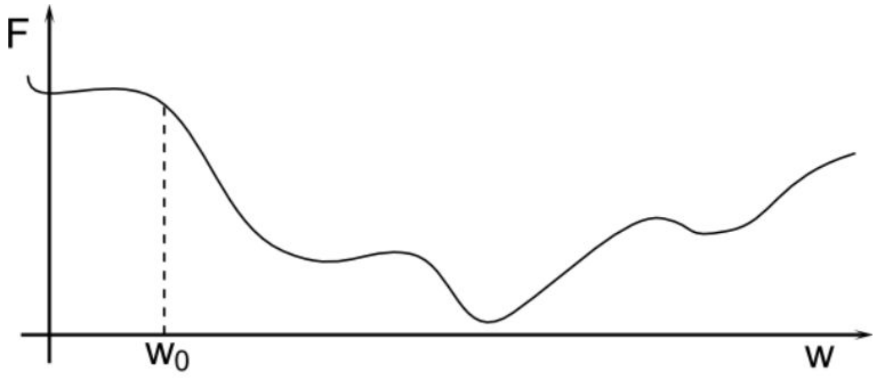
Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
    - Expectation Maximization (EM)
    - Convex Concave Procedure (CCP)



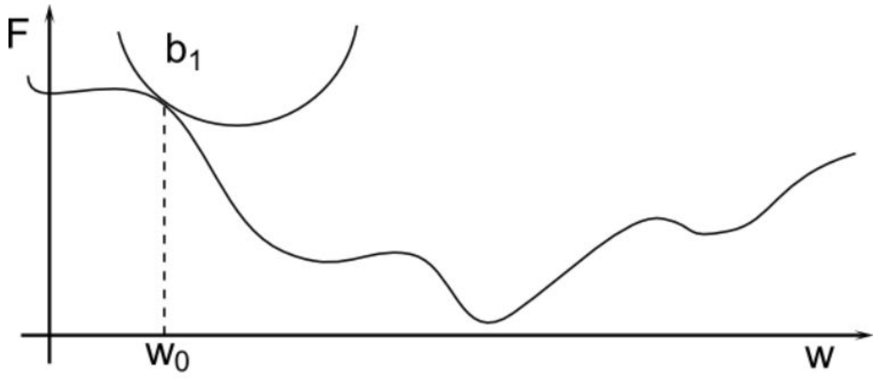
Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)



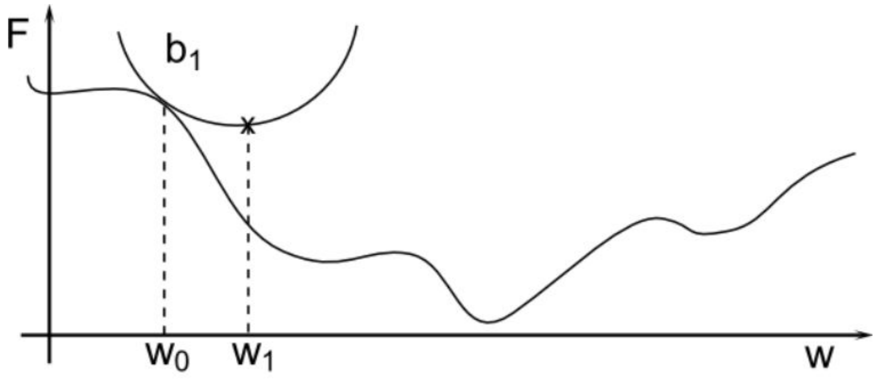
Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)



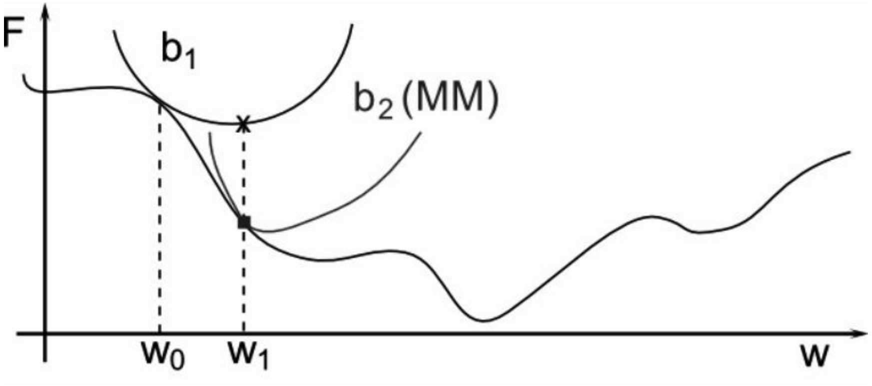
Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)



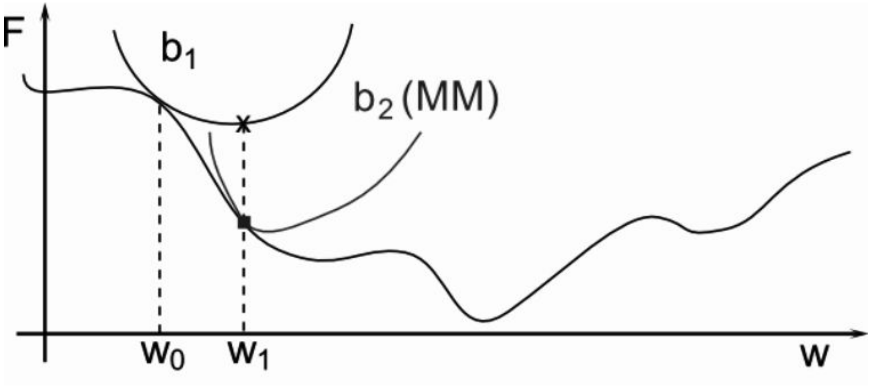
Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
    - Expectation Maximization (EM)
    - Convex Concave Procedure (CCP)



Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
    - Expectation Maximization (EM)
    - Convex Concave Procedure (CCP)

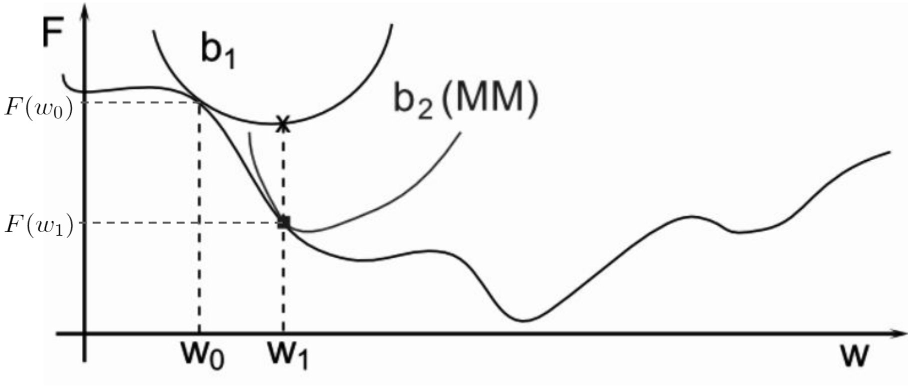


MM constraint:

$$b_t(w_{t-1}) = F(w_{t-1})$$

Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)



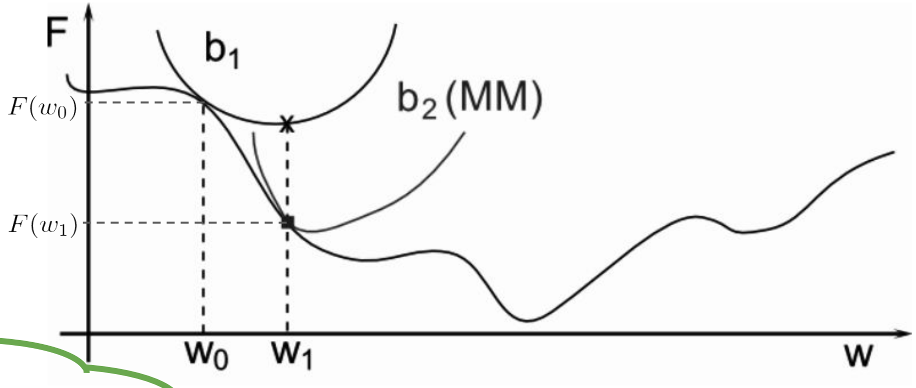
MM constraint:

$$b_t(w_{t-1}) = F(w_{t-1}) \qquad F(w_0) \geq F(w_1) \geq \cdots \geq F(w_T)$$

$\{F(w_t)\}_{t=0}^{T}$ Non-increasing sequence

Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)



Is this touching constraint necessary?

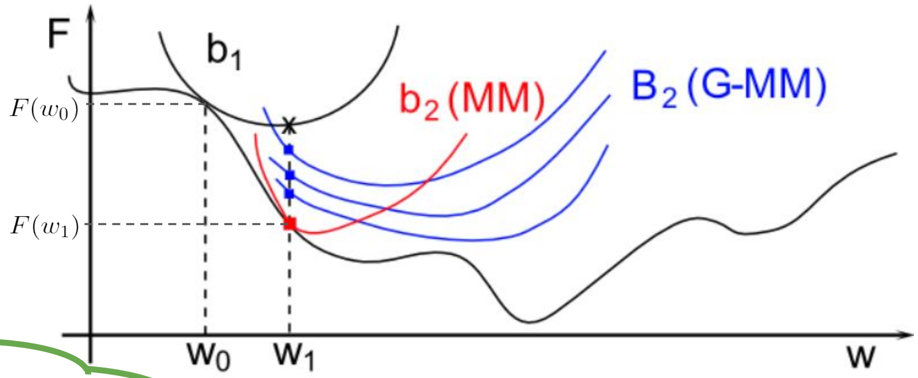
MM constraint:

$$b_t(w_{t-1}) = F(w_{t-1}) \qquad F(w_0) \geq F(w_1) \geq \cdots \geq F(w_T)$$

$\{F(w_t)\}_{t=0}^{T}$ Non-increasing sequence

Majorization Minimization (or Minorization Maximization)

- An iterative framework for non-convex optimization
- Examples of MM algorithm:
  - Expectation Maximization (EM)
  - Convex Concave Procedure (CCP)



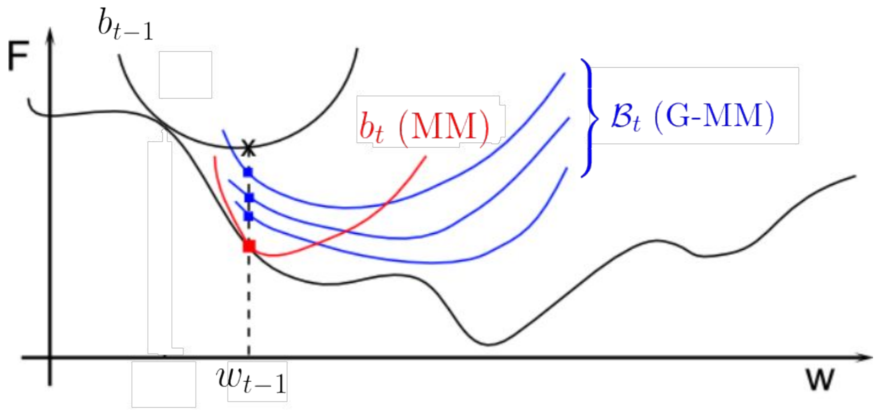
Is this touching constraint necessary?

MM constraint:

$$b_t(w_{t-1}) = F(w_{t-1}) \qquad F(w_0) \geq F(w_1) \geq \cdots \geq F(w_T)$$

$\{F(w_t)\}_{t=0}^{T}$ Non-increasing sequence

Bound selection



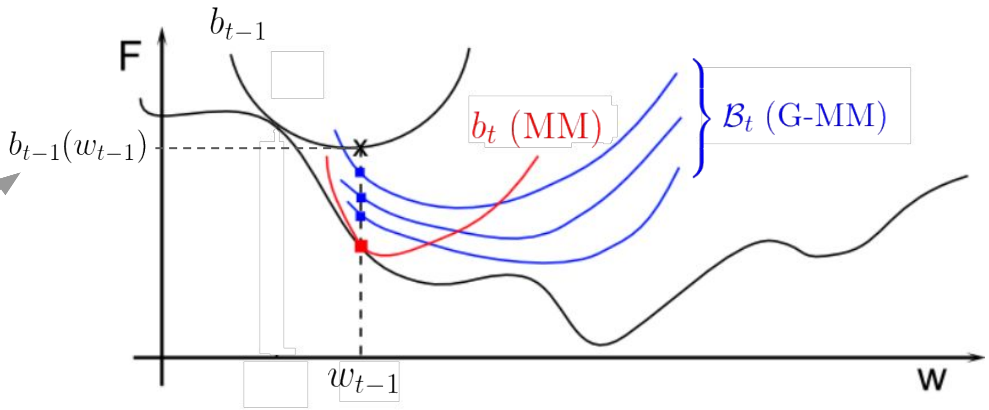
Bound selection

$\mathcal{F}$: family of bounds

$$\mathcal{B}(w, v) = \{b \in \mathcal{F} \mid b(w) \leq v\}$$

$$\mathcal{B}_t = \mathcal{B}(w_{t-1},\ b_{t-1}(w_{t-1}))$$

valid bounds
at iteration t



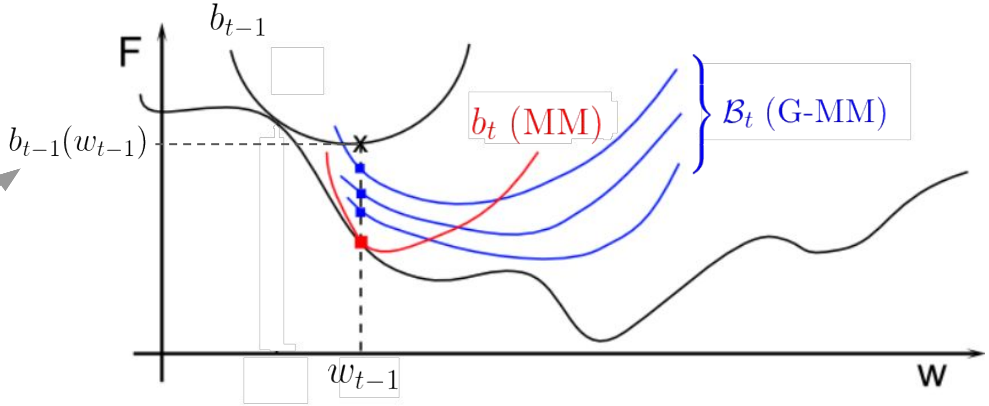
Bound selection

$\mathcal{F}$: family of bounds

$\mathcal{B}(w, v) = \{b \in \mathcal{F} \mid b(w) \leq v\}$

$\mathcal{B}_t = \mathcal{B}(w_{t-1},\ b_{t-1}(w_{t-1}))$

valid bounds at iteration t



Bound selection strategies:

- **Stochastic:** Sample uniformly from $\mathcal{B}_t$. $b_t \overset{U}{\sim} \mathcal{B}_t$

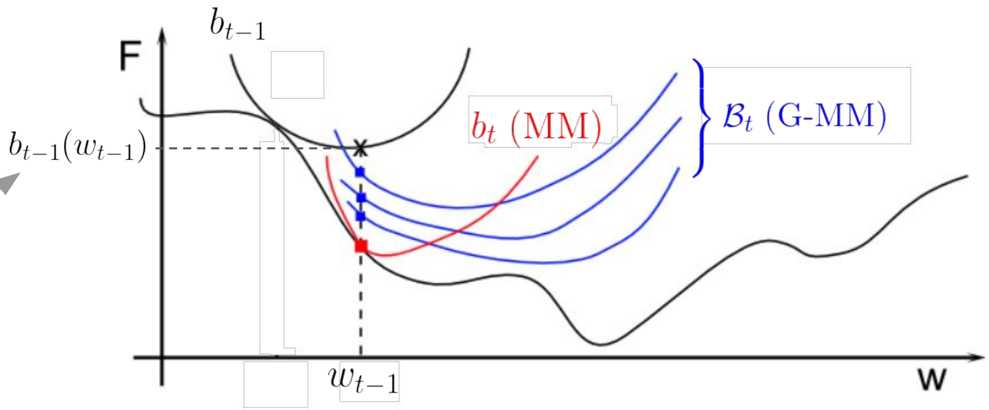
Bound selection

$\mathcal{F}$: family of bounds

$\mathcal{B}(w, v) = \{b \in \mathcal{F} \mid b(w) \leq v\}$

$\mathcal{B}_t = \mathcal{B}(w_{t-1},\ b_{t-1}(w_{t-1}))$

valid bounds
at iteration t



Bound selection strategies:

- **Stochastic:** Sample uniformly from $\mathcal{B}_t$. $b_t \overset{U}{\sim} \mathcal{B}_t$
- **Deterministic:** Maximize a “score” function $g : \mathcal{B}_t \times \mathbb{R}^d \to \mathbb{R}$. $b_t = \arg\max_{b \in \mathcal{B}_t} g(b, w_{t-1})$

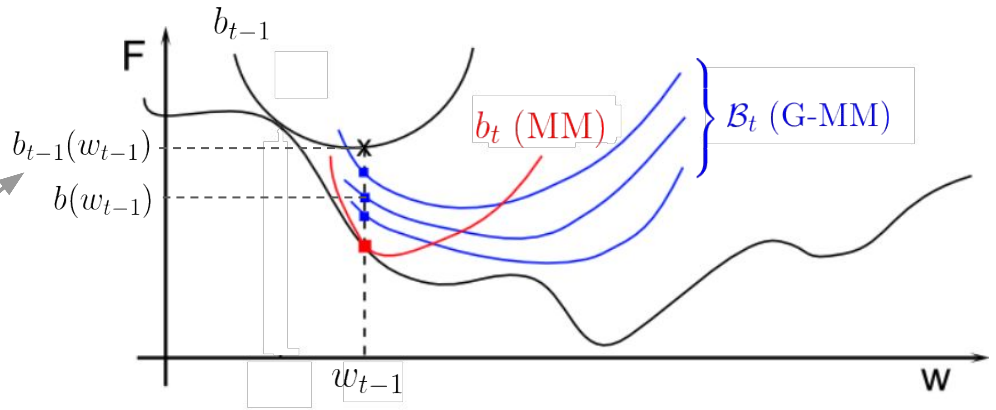
Bound selection

$\mathcal{F}$: family of bounds

$\mathcal{B}(w, v) = \{b \in \mathcal{F} \mid b(w) \leq v\}$

$\mathcal{B}_t = \mathcal{B}(w_{t-1},\ b_{t-1}(w_{t-1}))$

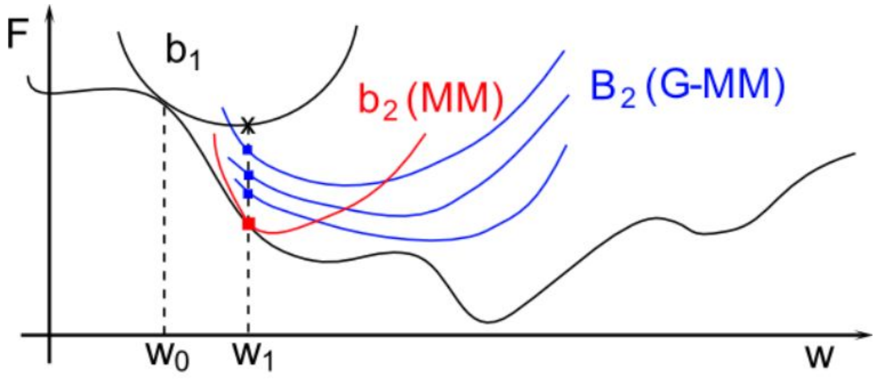
valid bounds
at iteration t



Bound selection strategies:

- **Stochastic:** Sample uniformly from $\mathcal{B}_t$. $b_t \overset{U}{\sim} \mathcal{B}_t$
- **Deterministic:** Maximize a "score" function $g : \mathcal{B}_t \times \mathbb{R}^d \to \mathbb{R}$. $b_t = \underset{b \in \mathcal{B}_t}{\arg\max}\ g(b, w_{t-1})$
  - E.g. MM corresponds to $g(b, w_{t-1}) = -b(w_{t-1})$.

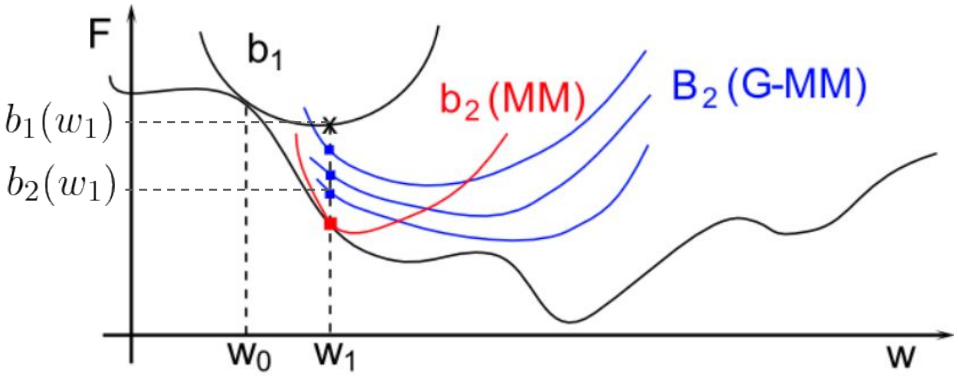
Generalized Majorization Minimization (G-MM)



G-MM constraint:

$$b_{t-1}(w_{t-1}) \geq b_t(w_{t-1})$$

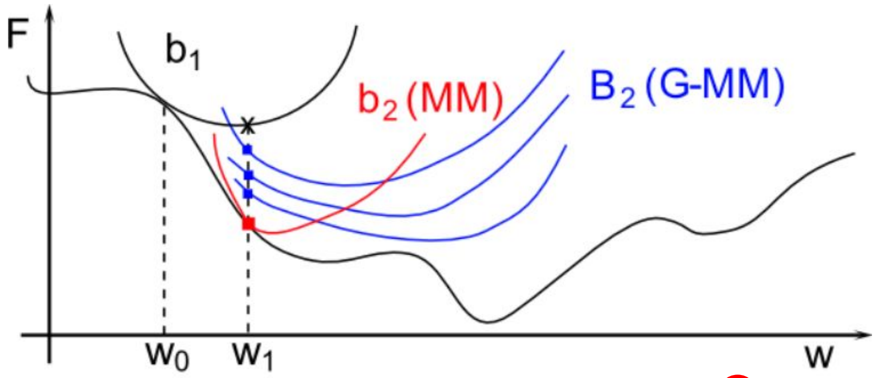
Generalized Majorization Minimization (G-MM)



G-MM constraint:

$$b_{t-1}(w_{t-1}) \geq b_t(w_{t-1})$$

Generalized Majorization Minimization (G-MM)



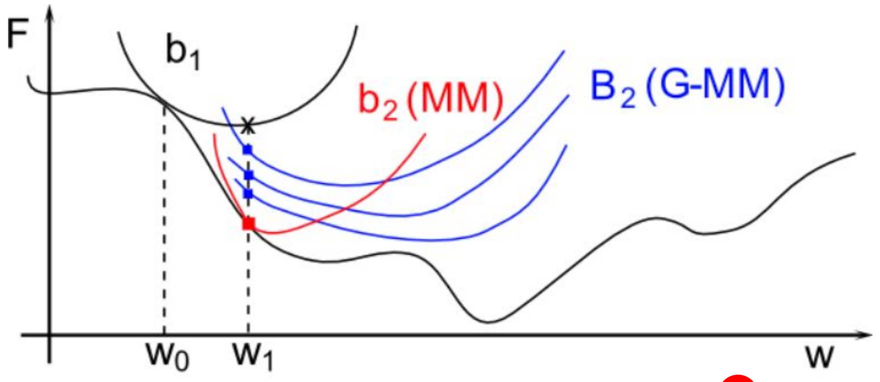
$\{b_t(w_t)\}_{t=1}^{T}$ Non-increasing sequence

G-MM constraint:

$$b_{t-1}(w_{t-1}) \geq b_t(w_{t-1})$$

$$b_1(w_1) \geq b_2(w_2) \geq \cdots \geq b_T(w_T)$$

Generalized Majorization Minimization (G-MM)



Theorem 1: $\lim_{T\to\infty} b_T(w_T) = F(w_T)$

Theorem 2: $\lim_{T\to\infty} \nabla F(w_T) = \mathbf{0}$

$\{b_t(w_t)\}_{t=1}^{T}$ Non-increasing sequence

G-MM constraint:

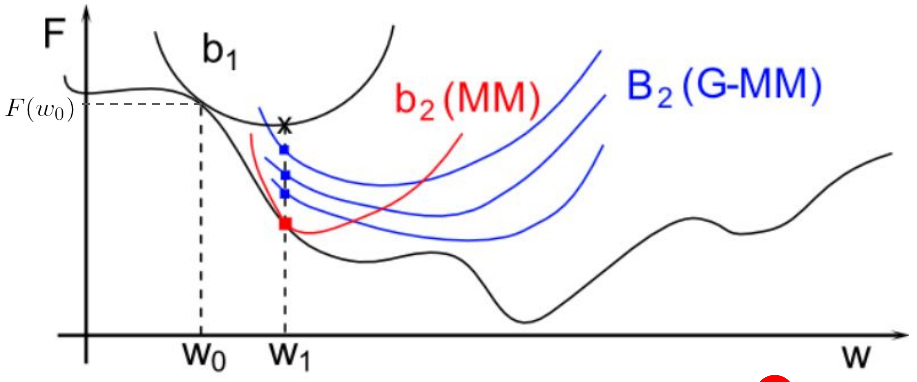
$$b_{t-1}(w_{t-1}) \geq b_t(w_{t-1})$$

$$b_1(w_1) \geq b_2(w_2) \geq \cdots \geq b_T(w_T) \;\rightarrow F(w_T)$$

Generalized Majorization Minimization (G-MM)

Theorem 1: $\lim_{T\to\infty} b_T(w_T) = F(w_T)$

Theorem 2: $\lim_{T\to\infty} \nabla F(w_T) = \mathbf{0}$



$F(w_0) \geq \{b_t(w_t)\}_{t=1}^{T}$ Non-increasing sequence

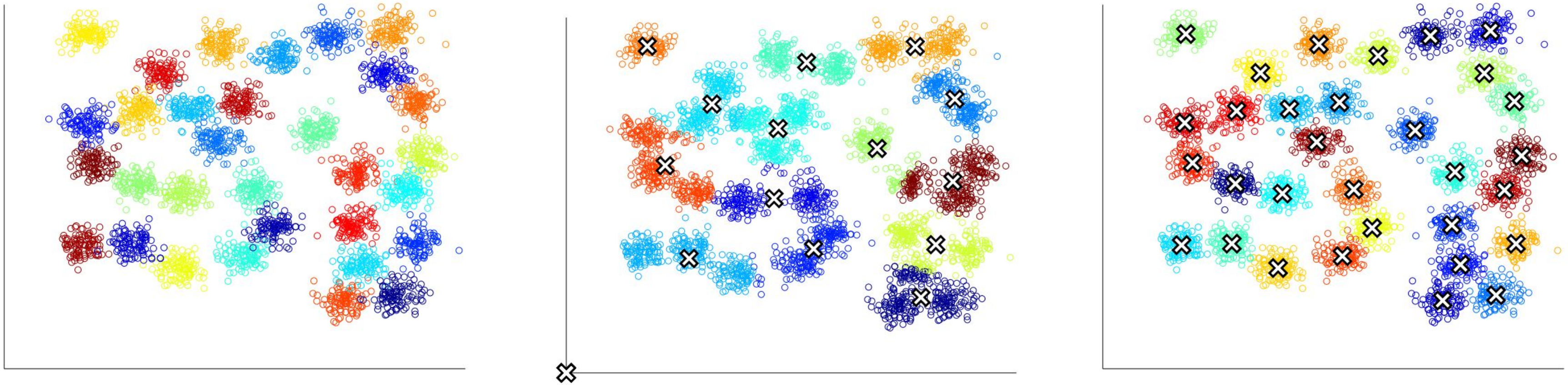
G-MM constraint:

$$b_{t-1}(w_{t-1}) \geq b_t(w_{t-1})$$

$$b_1(w_0) = F(w_0)$$

$$F(w_0) \geq b_1(w_1) \geq b_2(w_2) \geq \cdots \geq b_T(w_T) \rightarrow F(w_T)$$

G-MM: Results on clustering



(a) ground truth (b) k-means (c) generalized-MM

Qualitative analysis of the solutions found by MM (figure b) and G-MM (figure c).

Summary

- We proposed G-MM, an iterative optimization framework that generalizes MM.
- MM requires bounds to ***touch*** the objective function, which leads to sensitivity to initialization.
- We show that this touching constraint is unnecessary and relax it in G-MM.
- MM measures progress w.r.t. objective values → $(F(w_t))_{t=0}^{T}$ is non-increasing.
- G-MM measures progress w.r.t. bound values → $(b_t(w_t))_{t=1}^{T}$ is non-increasing.

$$w_t = \arg\min_{w} b_t(w)$$

- In each iteration of G-MM, a new bound is chosen from a **set** of valid bounds $\mathcal{B}_t$.
- Our experimental results, on several non-convex optimization problems, show that …
  - G-MM is less sensitive to initialization.
  - G-MM converges to solutions that have better objective value and perform better on the task.
  - G-MM can inject randomness to the optimization framework by choosing $b_t \overset{U}{\sim} \mathcal{B}_t$.
  - G-MM can incorporate biases into the optimization framework by choosing $b_t = \underset{b \in \mathcal{B}_t}{\arg\max}\ g(b, w_{t-1})$.