Optimizing Recommendation Systems with the Frank-Wolfe Algorithm

Optimization for Data Science

Anna Putina[1]

[1]Università degli Studi di Padova

February 5th, 2025

Università degli Studi di Padova

Table of Contents

Università degli Studi di Padova

Table of Contents

Università degli Studi di Padova

Frank-Wolfe Algorithm

Algorithm 1: Frank-Wolfe Method

Input: Convex function $f(x)$, feasible region C, initial point $x_0 \in C$

for $k = 0, 1, \ldots$ **do**

Compute $s_k = \arg\min_{s \in C} \langle \nabla f(x_k), s \rangle$ using LMO;

Set direction $d_k^{FW} = s_k - x_k$;

Choose step size $\alpha_k \in (0, 1]$;

Update $x_{k+1} = x_k + \alpha_k d_k^{FW}$;

if *stopping condition is met* **then**

break;

UNIVERSITÀ DEGLI STUDI DI PADOVA

Table of Contents

Università degli Studi di Padova

Matrix Completion Problem

Objective: Retrieve a low-rank matrix $X \in \mathbb{R}^{n_1 \times n_2}$ from a sparse set of observed matrix entries $\{U_{ij}\}_{(i,j)\in J}$, where $J \subset [1:n_1] \times [1:n_2]$.

Formulation:

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} \; f(X) := \sum_{(i,j)\in J} (X_{ij} - U_{ij})^2$$

subject to $\operatorname{rank}(X) \leq \delta$.

Relaxation:

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} \sum_{(i,j)\in J} (X_{ij} - U_{ij})^2$$

subject to $\|X\|_* \leq \delta$.

Solution Approach

Convex Optimization Formulation:

$$C = \{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_* \leq \delta\} = \text{conv}\{\delta \mathbf{u}\mathbf{v}^T : \mathbf{u} \in \mathbb{R}^{n_1}, \mathbf{v} \in \mathbb{R}^{n_2}, \|\mathbf{u}\| = \|\mathbf{v}\| = 1\}.$$

Linear Minimization Oracle (LMO):

$$\text{LMO}_C(\nabla f(X_k)) \in \arg \min_{\|X\|_* \leq \delta} \text{tr}(\nabla f(X_k)^T X).$$

Implementation:

- Compute the gradient $\nabla f(X_k)$.
- Find the rank-one matrix $\delta \mathbf{u}_1 \mathbf{v}_1^T$, where $\mathbf{u}_1, \mathbf{v}_1$ are the left and right singular vectors of $-\nabla f(X_k)$ corresponding to its top singular value.
- Approximate X as a sparse combination of rank-one matrices.

Table of Contents

Università degli Studi di Padova

Datasets Overview

MovieLens 20M Dataset

- A widely recognized benchmark in recommendation systems research.
- Contains **20 million** ratings from **138,000** users for over **27,000** movies.
- Ratings range from **0.5 to 5.0**, with metadata (titles, genres, timestamps).

Anime User Score Dataset

- Focuses on the anime domain with **16,500** anime titles.
- Contains ratings from **270,033** users.
- Ratings range from **1 to 10**.
- Presents challenges such as **diverse user base** and **varying popularity**.

Università degli Studi di Padova

Table of Contents

Università degli Studi di Padova

Algorithm Tracking and Stopping Criteria

The algorithm monitors key metrics during execution:

- **Loss Function:** Measures squared error over observed entries:

$$f(X) = \sum_{(i,j)\in J} (X_{ij} - U_{ij})^2.$$

- **Dual Gap:** Indicates sub-optimality:

$$\text{Dual Gap} = \langle \nabla f(X_k), X_k - S_k \rangle,$$

  where S_k is the solution from the LMO.

- **Relative Change Between Iterates:**

$$\frac{\|X_{k+1} - X_k\|_F}{\|X_k\|_F}.$$

Stopping thresholds balance accuracy and efficiency.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Learning Rate Strategies

Several learning rate strategies determine the step size α_k at each iteration:

- **Diminishing Step Size:** Standard for Frank-Wolfe:

$$\alpha_k = \frac{2}{k+2}.$$

- **Exact Line Search:** Minimizes the loss function:

$$\alpha_k = \arg\min_{\alpha \in [0,1]} f(X_k + \alpha(S_k - X_k)).$$

- **Armijo Rule:** Ensures sufficient descent:

$$f(X_k + \alpha_k d_k) \leq f(X_k) + c\alpha_k \langle \nabla f(X_k), d_k \rangle,$$

where $d_k = S_k - X_k$ and $c \in (0,1)$ is a constant.

Performance Tracking

The algorithm's performance was evaluated using:

- **Loss Function:** Tracks error over iterations.
- **Dual Gap:** Monitors convergence to optimality.

Key results:

- Comparison of loss and dual gap for different learning rate strategies.
- Showcases the effectiveness and efficiency of the Frank-Wolfe algorithm for matrix completion.

Università degli Studi di Padova

Table of Contents

Università degli Studi di Padova

Convergence Results

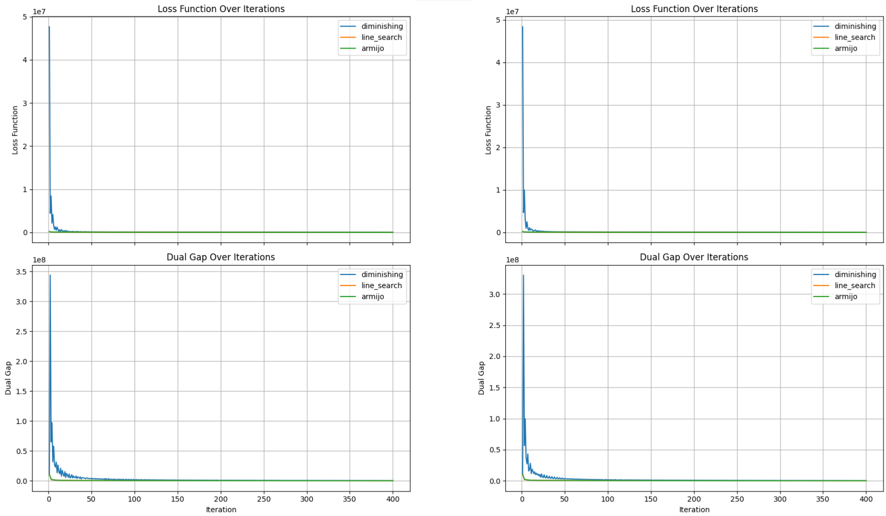

Figure 1. Comparison of loss and dual gap for different step sizes. Left: MovieLens dataset. Right: Anime dataset.

Università degli Studi di Padova

MovieLens

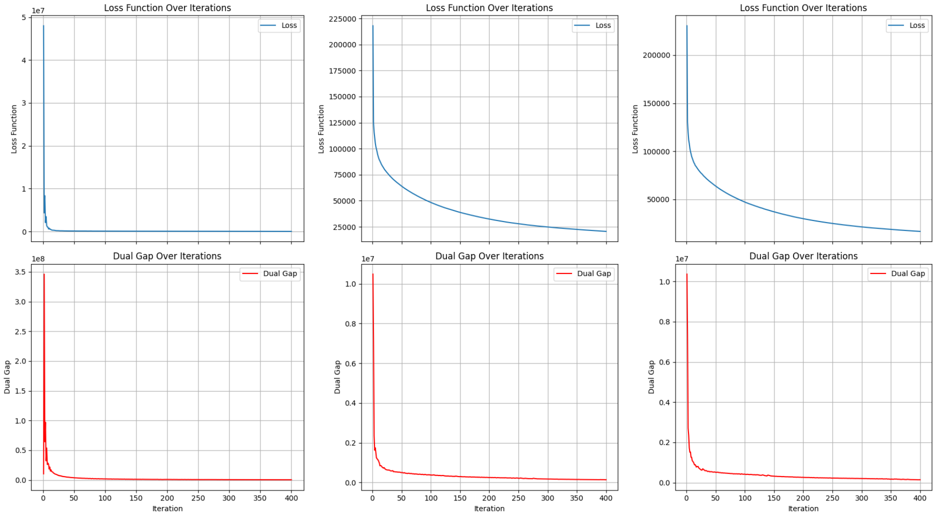

Figure 2. MovieLens dataset with different step size strategies: Diminishing, Exact Line Search, and Armijo Rule.

Anime

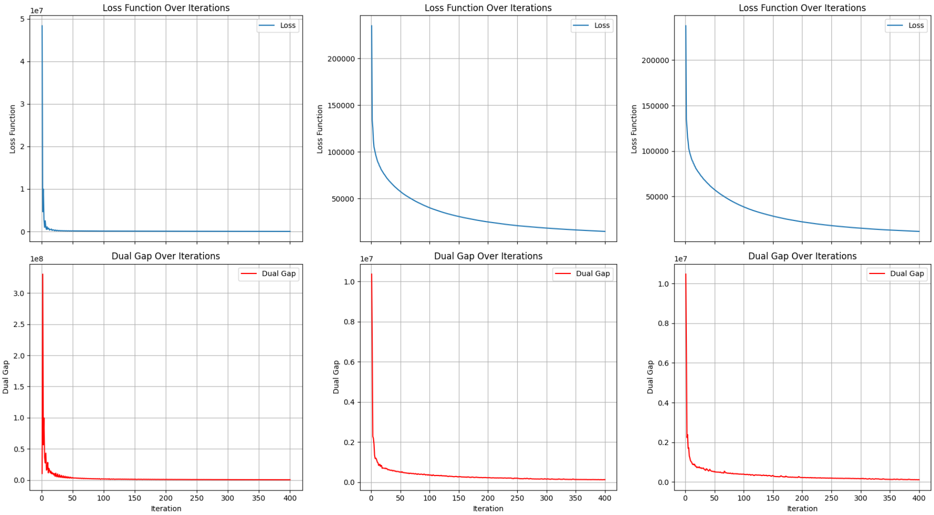

Figure 3. Anime dataset with different step size strategies: Diminishing, Exact Line Search, and Armijo Rule.

Università degli Studi di Padova

Relative Accuracy Function and Results

Function Description: The function computes the percentage of predictions (y_{pred}) that fall within a given **tolerance** of the true values (y_{true}).

Results:

Dataset	**Step Size Strategy**	**Relative Accuracy (%)**
MovieLens	Diminishing	62
MovieLens	Line Search	82
MovieLens	Armijo	88
Anime	Diminishing	82
Anime	Line Search	98
Anime	Armijo	99

Università degli Studi di Padova