Estimation of the Click Volume by Large Scale Regression Analysis

Yury Lifshits[1,2] and Dirk Nowotka[3]

[1]Steklov Institute of Mathematics at St.Petersburg,
[2]California Institute of Technology,
[3]University of Stuttgart

General area:

Fast web-scale algorithms for machine learning

General area:

Fast web-scale algorithms for machine learning

Least Squares: Find α minimizing $\|\alpha M - y\|_2$

General area:

Fast web-scale algorithms for machine learning

Least Squares: Find α minimizing $\|\alpha M - y\|_2$

Assume M has $k \ll mn$ nonzero elements

Can we solve least squares in $\mathcal{O}(k)$ time?

General area:

Fast web-scale algorithms for machine learning

Least Squares: Find α minimizing $\|\alpha M - y\|_2$

Assume M has $k \ll mn$ nonzero elements
Can we solve least squares in $\mathcal{O}(k)$ time?

Our contribution:

1. Solving least squares on sparse matrices
2. Application to on-line advertisement

Part I

Fast Algorithm for Solving Least Squares on Sparse Matrices

Geometric View on Least Squares

Find α minimizing $\|\alpha M - y\|_2$

Let V be a linear hull of M's rows $T_1, \ldots, T_m$

Geometric View on Least Squares

Find α minimizing $\|\alpha M - y\|_2$

Let V be a linear hull of M's rows $T_1, \ldots, T_m$

Interpretation: every vector α represents coordinates of αM point in V in $T_1, \ldots, T_m$ basis

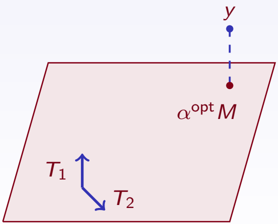
Geometric View on Least Squares

Find α minimizing $\|\alpha M - y\|_2$

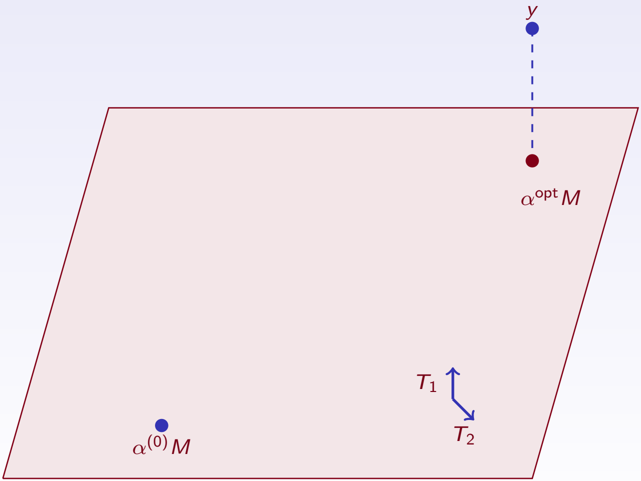
Let V be a linear hull of M's rows $T_1, \ldots, T_m$

Interpretation: every vector α represents coordinates of αM point in V in $T_1, \ldots, T_m$ basis

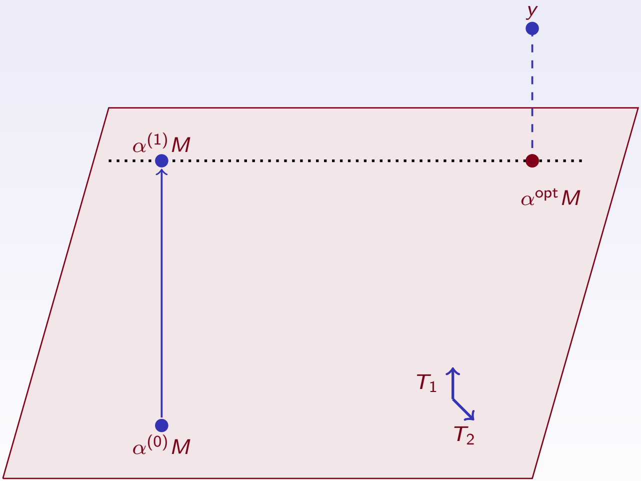
Thus $\alpha^{\text{opt}} M$ is just the projection of y to V



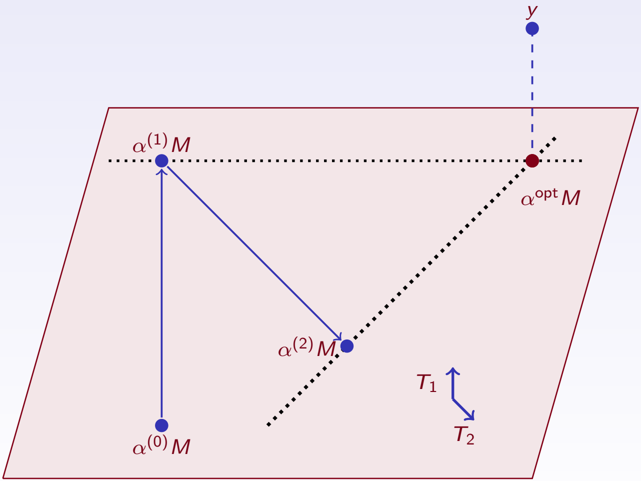
Componentwise Iterations: Idea



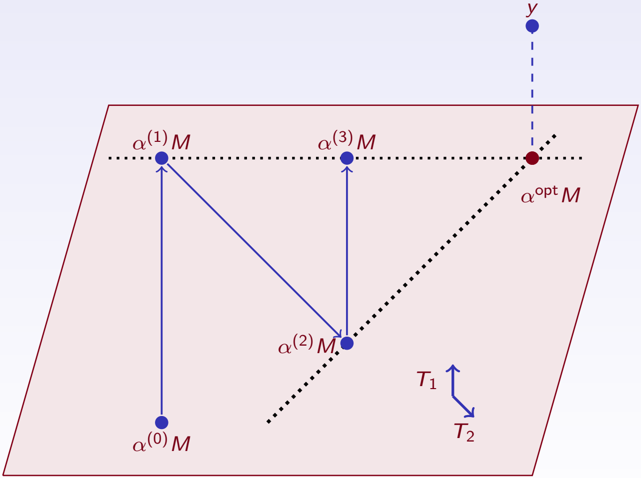
Componentwise Iterations: Idea



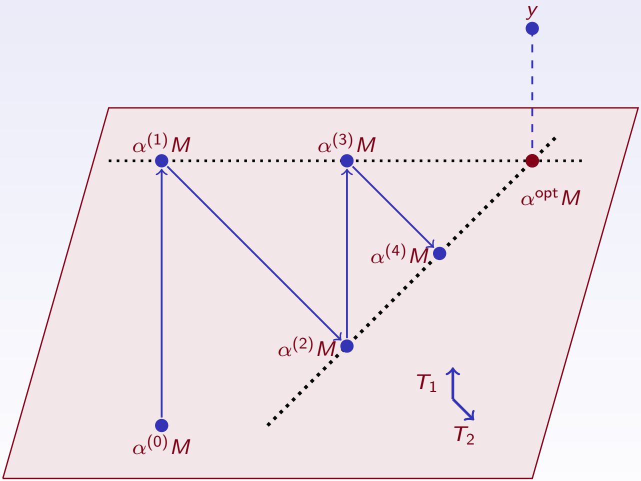
Componentwise Iterations: Idea



Componentwise Iterations: Idea



Componentwise Iterations: Idea



Componentwise Iterations: Algorithm

Auxiliary data structures:

Matrix columns: $T_1, \ldots, T_m$, precomputed norms $\|T_j\|$

Current solution $\alpha^{(k)}$

Error vector for current prediction $\gamma^{(k)} = y - \alpha^{(k)} M$

Componentwise Iterations: Algorithm

Auxiliary data structures:

Matrix columns: $T_1, \ldots, T_m$, precomputed norms $\|T_j\|$
Current solution $\alpha^{(k)}$
Error vector for current prediction $\gamma^{(k)} = y - \alpha^{(k)} M$

Componentwise iterations:

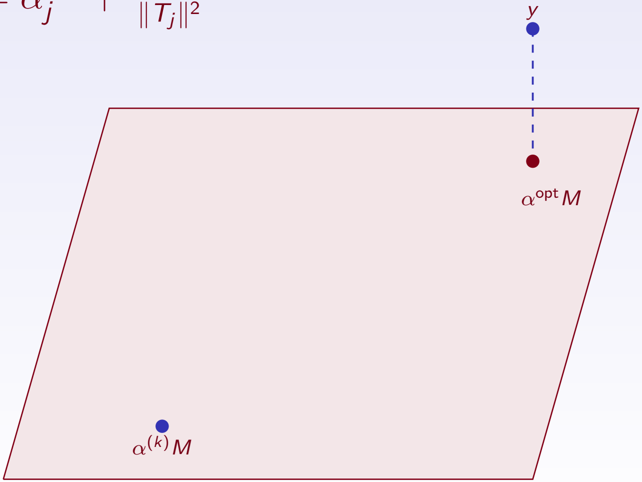
1. Start with random $\alpha^{(0)}$
2. Choose some coordinate j. Update formulae:

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} + \frac{\gamma^{(k)} \cdot T_j}{\|T_j\|^2}$$

$$\gamma^{(k+1)} = \gamma^{(k)} - \frac{\gamma^{(k)} \cdot T_j}{\|T_j\|^2} T_j$$

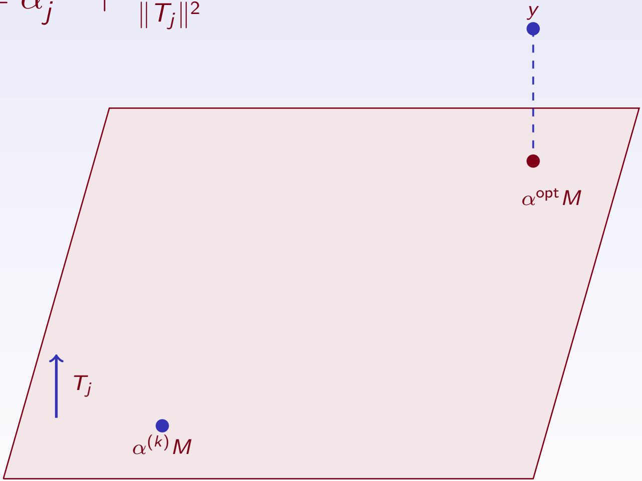
Optimality of One Step Update

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} + \frac{\gamma^{(k)} \cdot T_j}{\|T_j\|^2}$$



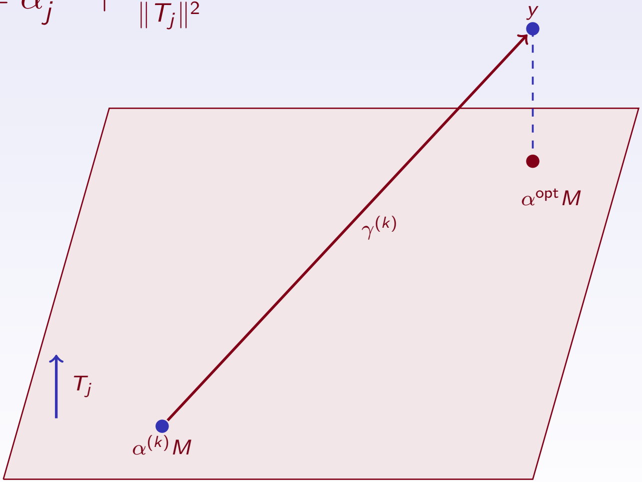
Optimality of One Step Update

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} + \frac{\gamma^{(k)} \cdot T_j}{\|T_j\|^2}$$



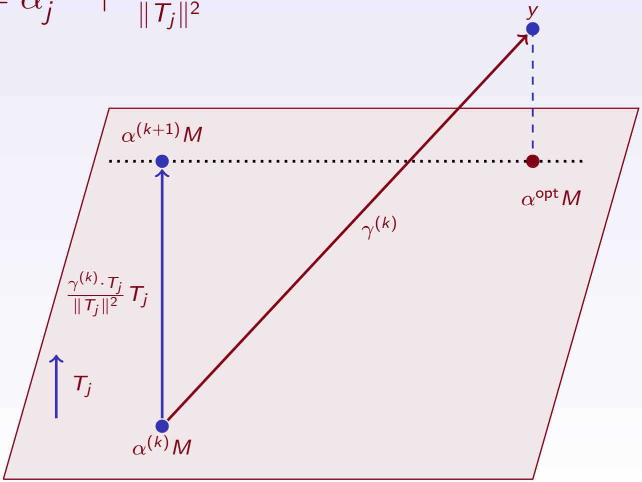
Optimality of One Step Update

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} + \frac{\gamma^{(k)} \cdot T_j}{\|T_j\|^2}$$



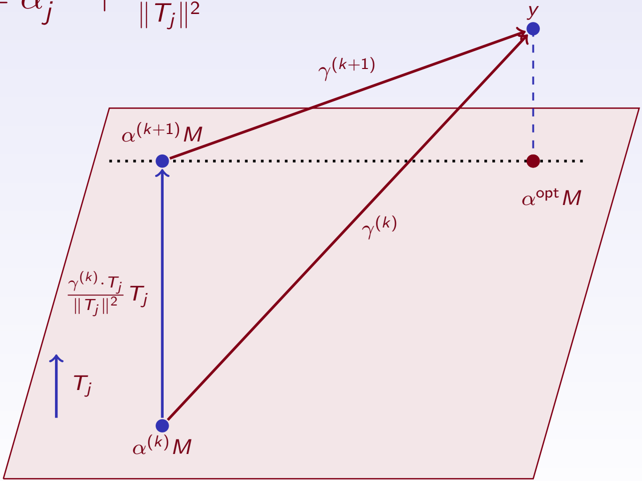
Optimality of One Step Update

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} + \frac{\gamma^{(k)} \cdot T_j}{\|T_j\|^2}$$



Optimality of One Step Update

$$\alpha_j^{(k+1)} = \alpha_j^{(k)} + \frac{\gamma^{(k)} \cdot T_j}{\| T_j \|^2}$$



Convergence Theorem

Theorem (Convergence theorem)

$\alpha^{(k)}$ converges to $\alpha^{opt} = (M \cdot M^T)^{-1} M^T y$ for any infinite order of updates containing every j infinitely many times

Convergence Theorem

Theorem (Convergence theorem)

$\alpha^{(k)}$ *converges to* $\alpha^{opt} = (M \cdot M^T)^{-1} M^T y$ *for any infinite order of updates containing every* j *infinitely many times*

Open Problem: prove some upper bounds on convergence speed

Complexity of One Global Round

Global Round:

Sequentially do one update for every j from 1 to m

Theorem

Componentwise Iterations algorithm uses only $\mathcal{O}(k)$ time for global round of updates. Recall, k is the number of nonzero entries in M.

Algorithm: Discussion

- A vector $\alpha^{(k)}$ can be safely updated in two components j_1 and j_2 *in parallel* if we have $T_{j_1} \perp T_{j_2}$
- In the case of orthogonal vectors $T_1, \ldots, T_m$ one global round is sufficient for reaching α^{opt}
- Joint optimal update of k components requires inverting $k \times k$ matrix

Part II

Application to On-line Advertisements

On-line Ads: Terminology

Opportunity: all available information about ad request to advertising system. E.g. website, search query, user id, . . .

On-line Ads: Terminology

Opportunity: all available information about ad request to advertising system. E.g. website, search query, user id, . . .

Click-through rate $CTR(a, o)$: a probability of clicking ad a at opportunity o

On-line Ads: Terminology

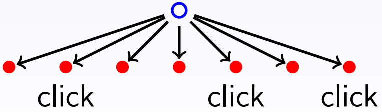
Opportunity: all available information about ad request to advertising system. E.g. website, search query, user id, ...

Click-through rate $CTR(a, o)$: a probability of clicking ad a at opportunity o

Click volume for collection O of all opportunities for some time interval: $CV(a) = \sum_{o \in O} CTR(a, o)$

The same ad

to all ad requests



Importance of Click Volume

- Measuring effectiveness of advertising system by a ration between sold clicks and click volume

Importance of Click Volume

- Measuring effectiveness of advertising system by a ration between sold clicks and click volume
- Measuring preciseness of targeting by ration between click volume in target group and overall click volume

Importance of Click Volume

- Measuring effectiveness of advertising system by a ration between sold clicks and click volume
- Measuring preciseness of targeting by ration between click volume in target group and overall click volume
- Defining coverage/agressiveness tradeoff: How many users should we contact in order to get a given fraction of click volume?

Importance of Click Volume

- Measuring effectiveness of advertising system by a ration between sold clicks and click volume
- Measuring preciseness of targeting by ration between click volume in target group and overall click volume
- Defining coverage/agressiveness tradeoff: How many users should we contact in order to get a given fraction of click volume?
- Helpful for setting prices for on-line ads

Estimation of Click Volume: Methodology

- Indexing: mapping advertising system logs to well-defined data structure
- Regression Analysis: deriving formula for click-through rate prediction
- Estimation: applying resulting formula to a given ad and monthly collection of opportunities

Indexing: History Table

Logs of advertising system:

Set of triples (a_i, o_i, b_i): ad, opportunity, click

Indexing: History Table

Logs of advertising system:

Set of triples (a_i, o_i, b_i): ad, opportunity, click

Events indexing:

Mapping a pair (a, o) to
internal representation $E(a, o)$
Techniques: clustering, dimensionality reduction

Indexing: History Table

Logs of advertising system:

Set of triples (a_i, o_i, b_i): ad, opportunity, click

Events indexing:

Mapping a pair (a, o) to
internal representation $E(a, o)$
Techniques: clustering, dimensionality reduction

Resulting reduced table:

Set of pairs (E_i, CTR_i)

Quick Recall: Regression Analysis

Input:

Training collection of n documents
Document i: m-dimensional vector D_i
and additional parameter y_i

Quick Recall: Regression Analysis

Input:

Training collection of n documents
Document i: m-dimensional vector D_i
and additional parameter y_i

Regression Problem:

Find function f such that $f(D_i) \approx y_i$

Quick Recall: Regression Analysis

Input:

Training collection of n documents
Document i: m-dimensional vector D_i
and additional parameter y_i

Regression Problem:

Find function f such that $f(D_i) \approx y_i$

Linear Regression Problem (least squares):

find m-dimensional vector α such that the sum of squared prediction errors $\sum |\alpha D_i - y_i|^2$ is minimized

Click Volume via Regression

1. Building history table from logs of advertising system: matrix $M = \{E_1, \ldots, E_n\}$ and CTR vector

Click Volume via Regression

1. Building history table from logs of advertising system: matrix $M = \{E_1, \ldots, E_n\}$ and CTR vector
2. Solving least squares

Click Volume via Regression

1. Building history table from logs of advertising system: matrix $M = \{E_1, \ldots, E_n\}$ and CTR vector
2. Solving least squares
3. Computing click volume by formula:

$$CV(a_{\text{new}}) = \sum_{1 \le i \le n} \alpha \cdot E(a_{\text{new}}, o_i)$$

Directions for Further Work

Theoretic problems:

- Prove upper bounds for speed of convergence of our algorithm
- Can we compute click volume for all ads in the system faster than doing it separately for every ad?

Directions for Further Work

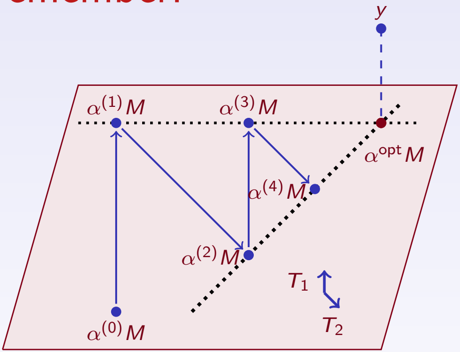
Theoretic problems:

- Prove upper bounds for speed of convergence of our algorithm
- Can we compute click volume for all ads in the system faster than doing it separately for every ad?

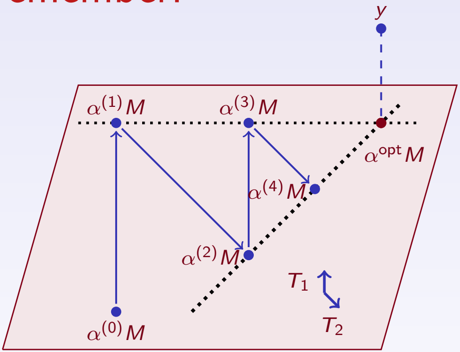
Experimental problems:

- Apply our algorithm for some real data set. Measure the empirical precision of *CTR* prediction
- Study effects of heuristical ingredients for algorithm: indexing, dimensionality reduction, update order

A picture to remember:



A picture to remember:



Thanks for your attention! Questions?

Yury Lifshits http://yury.name

Dirk Nowotka google://dirk nowotka

Some related work:

Y. Lifshits, D. Nowotka
Estimation of the Click Volume by Large Scale Regression Analysis. CSR'07.
http://yury.name/papers/lifshits2007click.pdf

B. Hoffmann, Y. Lifshits, D. Nowotka
Maximal Intersection Queries in Randomized Graph Models. CSR'07.
http://yury.name/papers/hoffmann2007maximal.pdf

N. Goyal, Y. Lifshits, H. Schütze
Disorder Inequality: A Combinatorial Approach to Nearest Neighbor Search. Submitted.
http://yury.name/papers/goyal2008disorder.pdf