Inexact Predictor-Corrector Methods for Linear Programming

Gregory Dexter
Department of Computer Science
Purdue University

PURDUE UNIVERSITY

Joint work with...

Agniva Chowdhury
Oak Ridge National Laboratory

Haim Avron
Tel Aviv University

Petros Drineas
Purdue University

Linear Programming (LP)

Consider the standard form of the primal LP problem:

$$\min \mathbf{c}^\mathsf{T}\mathbf{x}\,, \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b}\,, \mathbf{x} \geq \mathbf{0} \qquad (1)$$

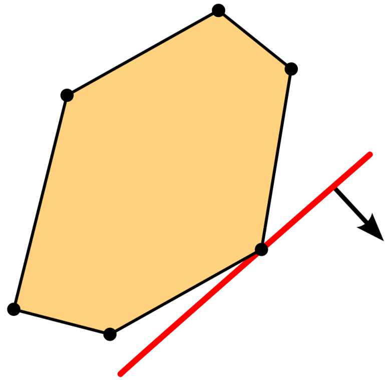
The associated dual problem is

$$\max \mathbf{b}^\mathsf{T}\mathbf{y}\,, \text{ subject to } \mathbf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} = \mathbf{c}\,, \mathbf{s} \geq \mathbf{0} \qquad (2)$$

Here,

$\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{c} \in \mathbb{R}^n$ are inputs

$\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, and $\mathbf{s} \in \mathbb{R}^n$ are variables



An LP problem with $m = 6, n = 2$.

LP: Applications in ML

- Basis pursuit [Tillmann , PAMM 2015]
- Sparse inverse covariance matrix estimation (SICE) [Yuan , JMLR 2010]
- MAP inference [Meshi & Globerson , ECML PKDD 2011]
- ℓ_1-regularized SVMs [Zhu, Rosset, Tibshirani, & Hastie , NeurIPS 2004]
- Nonnegative matrix factorization (NMF) [Recht et al. , NeurIPS 2012]
- Markov decision process (MDP) [Bello & Riano , IEEE SIEDS 2006]

Objective Overview

Goal: Speed up linear programming on large-scale data sets for "big data" applications, such as found in ML and computational biology

- Focus on using using practical algorithms, i.e.,
  - Predictor-corrector methods instead of short step
  - Iterative linear solvers instead of fast matrix multiplication
  - Efficient preconditioner construction instead of inverse maintenance
- Extend classic theoretical convergence guarantees for linear programming to allow for the use of inexact linear system solves

Optimality conditions

$(\mathbf{x}, \mathbf{y}, \mathbf{s})$ is an (primal-dual) optimal solution iff it satisfies the following conditions:[1]

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0} \qquad \text{(primal feasibility)}$$

$$\mathbf{A}^{\mathsf{T}}\mathbf{y} + \mathbf{s} = \mathbf{c}, \ \mathbf{s} \geq \mathbf{0} \qquad \text{(dual feasibility)}$$

$$\mathbf{x} \circ \mathbf{s} = \mathbf{0} \qquad \text{(complementary slackness)}$$

Assumptions:

- $n \gg m$ and $\mathrm{rank}(\mathbf{A}) = m$
- Solution set is nonempty

[1] Let $\mathbf{x} \circ \mathbf{s}$ denote the entry-wise product of $\mathbf{x}$ and $\mathbf{s}$, i.e., $[\mathbf{x} \circ \mathbf{s}]_i = \mathbf{x}_i \mathbf{s}_i$

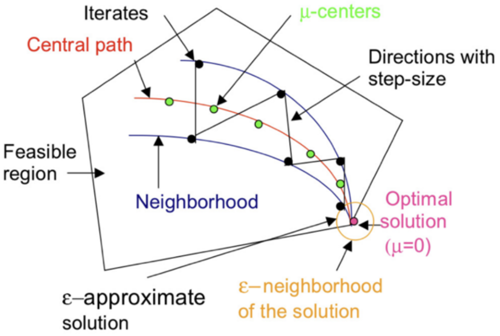
Standard Methods

Simplex

- Fast in practice
- $\exp$-time worst case

Interior Point

- Fastest in theory
- Often faster in practice for large-scale LPs



Path-following IPM visualization. Figure from [2].

Interior point methods

- **Duality measure**:

$$\mu = \frac{\mathbf{x}^\mathsf{T}\mathbf{s}}{n} = \frac{\mathbf{x}^\mathsf{T}(\mathbf{c} - \mathbf{A}^\mathsf{T}\mathbf{y})}{n} = \frac{\mathbf{c}^\mathsf{T}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{y}}{n} \downarrow 0$$

- **Feasible Predictor-Corrector IPM:**
  - Let $\mathcal{F}^0 = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) : (\mathbf{x}, \mathbf{s}) > \mathbf{0},\ \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} = \mathbf{c}\}$.

Interior point methods

- **Duality measure**:

$$\mu = \frac{\mathbf{x}^\mathsf{T}\mathbf{s}}{n} = \frac{\mathbf{x}^\mathsf{T}(\mathbf{c} - \mathbf{A}^\mathsf{T}\mathbf{y})}{n} = \frac{\mathbf{c}^\mathsf{T}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{y}}{n} \downarrow 0$$

- **Feasible Predictor-Corrector IPM:**
  - Let $\mathcal{F}^0 = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) : (\mathbf{x}, \mathbf{s}) > \mathbf{0},\ \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} = \mathbf{c}\}$.
  - Central path: $\mathcal{C} = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^0 : \mathbf{x} \circ \mathbf{s} = \mu \mathbf{1}_n\}$, where $\mathbf{x} \circ \mathbf{s}$ denotes the element-wise product of $\mathbf{x}$ and $\mathbf{s}$.

Interior point methods

- **Duality measure**:

$$\mu = \frac{\mathbf{x}^\mathsf{T}\mathbf{s}}{n} = \frac{\mathbf{x}^\mathsf{T}(\mathbf{c} - \mathbf{A}^\mathsf{T}\mathbf{y})}{n} = \frac{\mathbf{c}^\mathsf{T}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{y}}{n} \downarrow 0$$

- **Feasible Predictor-Corrector IPM:**
  - Let $\mathcal{F}^0 = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) : (\mathbf{x}, \mathbf{s}) > \mathbf{0},\ \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} = \mathbf{c}\}$.
  - Central path: $\mathcal{C} = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^0 : \mathbf{x} \circ \mathbf{s} = \mu \mathbf{1}_n\}$, where $\mathbf{x} \circ \mathbf{s}$ denotes the element-wise product of $\mathbf{x}$ and $\mathbf{s}$.
  - Neighborhood: $\mathcal{N}_2(\theta) = \left\{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^0 : \|\mathbf{x} \circ \mathbf{s} - \mu \mathbf{1}_n\|_2 \leq \theta\mu,\ (\mathbf{x}, \mathbf{s}) > \mathbf{0}\right\}$

Solving linear system

Let $\mathbf{X}$ and $\mathbf{S}$ be diagonal matrices with entries of $\mathbf{x}$ and $\mathbf{s}$ on the diagonal respectively.

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^\top & \mathbf{I}_n \\ \mathbf{S} & \mathbf{0} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \\ \Delta\mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{XS1}_n + \sigma\mu\mathbf{1}_n \end{pmatrix}$$

$$\mathbf{AD}^2\mathbf{A}^\top\Delta\mathbf{y} = \underbrace{-\sigma\mu\mathbf{AS}^{-1}\mathbf{1}_n + \mathbf{Ax}}_{\mathbf{p}}, \tag{3}$$

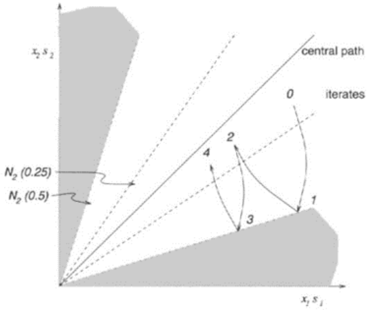
$$\Delta\mathbf{s} = -\mathbf{A}^\top\Delta\mathbf{y}, \tag{4}$$

$$\Delta\mathbf{x} = -\mathbf{x} + \sigma\mu\mathbf{S}^{-1}\mathbf{1}_n - \mathbf{D}^2\Delta\mathbf{s}. \tag{5}$$

Here, $\mathbf{D} = \mathbf{X}^{1/2}\mathbf{S}^{-1/2}$ is a diagonal matrix.

Predictor-Corrector Method

1. Start in the smaller neighborhood $\mathcal{N}_2(0.25)$
2. Take a *predictor step*
   - centering parameter $\sigma = 0$
   - Remains within the larger $\mathcal{N}_2(0.5)$ neighborhood
   - Makes large progress towards the optimum
3. Take a *corrector step*
   - centering parameter $\sigma = 1$
   - Goes towards the central path
   - Returns to the smaller $\mathcal{N}_2(0.25)$ neighborhood
4. Repeat until the duality measure μ is less than ϵ



Predictor-corrector visualization.
Figure from [3]

Solving normal equation

$$\mathbf{A}\mathbf{D}^2\mathbf{A}^\top \Delta \mathbf{y} = \mathbf{p} \tag{3}$$

Direct solvers

- If $\mathbf{A}$ is high-dimensional and dense, computationally prohibitive.
- Sparse solvers doesn't take into account the irregular sparsity pattern of $\mathbf{A}\mathbf{D}^2\mathbf{A}$.

Iterative solvers

- $\mathbf{A}\mathbf{D}^2\mathbf{A}^\top$ is typically ill-conditioned near the optimal solution.
- Does not return an exact solution (invalidates standard theoretical analysis)
- Does not maintain primal feasibility

Structural Condition: Inexact system solve

We can maintain $\mathcal{O}(\sqrt{n}\log\frac{\mu_0}{\epsilon})$ outer iteration complexity as long as an inexact solver satisfies at each iteration:[2]

$$\|\Delta\tilde{\mathbf{y}} - (\mathbf{A}\mathbf{D}^2\mathbf{A}^T)^{-1}\mathbf{p}\|_{\mathbf{A}\mathbf{D}^2\mathbf{A}^T} \le \delta \quad \text{and} \quad \|\mathbf{A}\mathbf{D}^2\mathbf{A}^T\Delta\tilde{\mathbf{y}} - \mathbf{p}\|_2 \le \delta,$$

$$\text{with } \delta = \mathcal{O}\left(\frac{\epsilon}{\sqrt{n}\log\mu_o/\epsilon}\right).$$

- Running the standard predictor-correct algorithm with such an inexact solver converges in $\mathcal{O}(\sqrt{n}\log\frac{\mu_0}{\epsilon})$ outer iterations to an ϵ-optimal solution (same as using a direct solver)
- The final solution will be ϵ-feasible, i.e., $\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2 \le \epsilon$.

[2] The *energy-norm* is denoted as $\|\mathbf{x}\|_{\mathbf{M}} = \sqrt{\mathbf{x}^T\mathbf{M}\mathbf{x}}$ for vector $\mathbf{x}$ and PSD matrix $\mathbf{M}$.

Structural Condition: Error-adjusted solver

How do we ensure that the final solution is <u>exactly</u> feasible?

Perturbation vector $\mathbf{v}$ [Monteiro and O'Neal, 2003]

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^\top & \mathbf{I}_n \\ \mathbf{S} & \mathbf{0} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \Delta\tilde{\mathbf{x}} \\ \Delta\tilde{\mathbf{y}} \\ \Delta\tilde{\mathbf{s}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{X}\mathbf{S}\mathbf{1}_n + \sigma\mu\mathbf{1}_n - \mathbf{v} \end{pmatrix}$$

$$\mathbf{A}\mathbf{D}^2\mathbf{A}^\top\Delta\tilde{\mathbf{y}} = \mathbf{p} + \mathbf{A}\mathbf{S}^{-1}\mathbf{v}\,, \tag{6}$$

$$\Delta\tilde{\mathbf{s}} = -\mathbf{A}^\top\Delta\tilde{\mathbf{y}}\,, \tag{7}$$

$$\Delta\tilde{\mathbf{x}} = -\mathbf{x} + \sigma\mu\mathbf{S}^{-1}\mathbf{1}_n - \mathbf{D}^2\Delta\tilde{\mathbf{s}} - \mathbf{S}^{-1}\mathbf{v}. \tag{8}$$

- $\mathbf{A}\Delta\tilde{\mathbf{x}} = \mathbf{0}$ if $\mathbf{v}$ satisfies eqn. (6) $\Rightarrow$ $\mathbf{A}(\mathbf{x} + \alpha\Delta\tilde{\mathbf{x}}) = \mathbf{b}$

Structural Condition: Error-adjusted solver

As long as the returned inexactly solution $\Delta\tilde{\mathbf{y}}$ and correction vector $\mathbf{v}$ satisfy:

$$\mathbf{A}\mathbf{D}^2\mathbf{A}^T\Delta\tilde{\mathbf{y}} = \mathbf{p} + \mathbf{A}\mathbf{S}^{-1}\mathbf{v} \quad \text{and} \quad \|\mathbf{v}\|_2 < \mathcal{O}(\epsilon), \tag{9}$$

- The modified predictor-corrector algorithm converges in $\mathcal{O}\left(\sqrt{n}\log\frac{\mu_0}{\epsilon}\right)$ outer iterations
- The final solution will be exactly feasible, i.e., $\mathbf{A}\mathbf{x}^* = \mathbf{b}$.

Iterative Solver

How can we efficiently solve the linear systems while fulfilling the previous structural conditions?

Iterative solver

Preconditioned Gradient Algorithm (PCG):[3]
Input: $\mathbf{AD} \in \mathbb{R}^{m\times n}$ with $m \ll n, \mathbf{p} \in \mathbb{R}^m$, sketching matrix $\mathbf{W} \in \mathbb{R}^{n\times w}$, iteration count t;

Step 1. Compute $\mathbf{ADW}$ and its SVD. Let $\mathbf{U_Q} \in \mathbb{R}^{m\times m}$ be the matrix of its left singular vectors and let $\mathbf{\Sigma}_{\mathbf{Q}}^{1/2} \in \mathbb{R}^{m\times m}$ be the matrix of its singular values;

Step 2. Compute $\mathbf{Q}^{-1/2} = \mathbf{U_Q}\mathbf{\Sigma}_{\mathbf{Q}}^{-1/2}\mathbf{U}_{\mathbf{Q}}^{\top}$;

Step 3. Initialize $\tilde{\mathbf{z}}^0 \leftarrow \mathbf{0}_m$ and run standard CG on $\mathbf{Q}^{-1/2}\mathbf{AD}^2\mathbf{A}^T\mathbf{Q}^{-1/2}\tilde{\mathbf{z}} = \mathbf{Q}^{-1/2}\mathbf{p}$ for t iterations;

Output: return $\hat{\Delta\mathbf{y}} = \mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^t$

- Sketching matrix $\mathbf{W}$ is an ℓ_2-subspace embedding matrix
- Used to construct a strong preconditioner $\mathbf{Q}^{-1,2}$ to reduce the condition number of the system to a constant
- Iterative solvers, e.g. PCG, converge exponentially quickly via standard analysis:

$$\|\mathbf{Q}^{-1/2}(\mathbf{AD}^2\mathbf{A}^T)\mathbf{Q}^{-1/2}\tilde{\mathbf{z}}^t - \mathbf{Q}^{-1/2}\mathbf{p}\|_2$$
$$\leq \zeta^t\|\mathbf{Q}^{-1/2}\mathbf{p}\|_2, \text{ for some } \zeta \in (0,1).$$

[3] First proposed in [1].

Inexact system solver for unmodified PC

Recall that the normal equations must be solved to the following precision with $\delta = \mathcal{O}\left(\frac{\epsilon}{\sqrt{n}\log \mu_o/\epsilon}\right)$:

$$\|\Delta\tilde{\mathbf{y}} - (\mathbf{A}\mathbf{D}^2\mathbf{A}^T)^{-1}\mathbf{p}\|_{\mathbf{A}\mathbf{D}^2\mathbf{A}^T} \leq \delta \quad \text{and} \quad \|\mathbf{A}\mathbf{D}^2\mathbf{A}^T\Delta\tilde{\mathbf{y}} - \mathbf{p}\|_2 \leq \delta.$$

- The previous PCG method will satisfy both conditions after $\mathcal{O}\left(\log \frac{\sigma_{\max}(\mathbf{AD})\, n\mu}{\epsilon}\right)$ iterations.
- The $\sigma_{\max}(\mathbf{AD})$ factor is needed to satisfy the ℓ_2-norm guarantee on the residual

Inexact system solver for error-adjusted PC

Recall that the inexact solution to the normal equations, $\Delta\tilde{\mathbf{y}}$, and correction vector, $\mathbf{v}$, must satisfy:

$$\mathbf{AD}^2\mathbf{A}^T\Delta\tilde{\mathbf{y}} = \mathbf{p} + \mathbf{AS}^{-1}\mathbf{v} \quad \text{and} \quad \|\mathbf{v}\|_2 < \mathcal{O}(\epsilon).$$

- It suffice to run for the PCG method for $\mathcal{O}\left(\log\frac{n\mu}{\epsilon}\right)$ iterations
- Notice the lack of the $\sigma_{\max}(\mathbf{AD})$ factor.

Correction vector

$$\mathbf{v} = (\mathbf{XS})^{1/2}\mathbf{W}(\mathbf{ADW})^{\dagger}(\mathbf{AD}^2\mathbf{A}^{\mathsf{T}}\hat{\Delta\mathbf{y}} - \mathbf{p}).$$

- Computable with a constant number of mat-vecs with already computed matrices.

Inexact solve time complexity

For the PCG solver instantiation...

- The preconditioner $\mathbf{Q}^{-1/2}$ can be computed efficiently if $\mathbf{W}$ is the count sketch matrix
  - $\mathbf{Q}^{-1/2}$ can be computed in $\mathcal{O}\left(m^3 \log \frac{m}{\eta}\right)$ time with probability at least $1-\eta$
- Each iteration of CG computes a constant number of matrix products with $\mathbf{Q}^{-1/2}$, $\mathbf{AD}$, and $\mathbf{DA}^T$.
  - Each mat-vec takes $\mathcal{O}(\text{nnz}(\mathbf{A}) + m^3)$ time
- Total number of iterations is logarithmic in n
  - $\mathcal{O}\left(\log \frac{\sigma_{\max}(\mathbf{AD})\, n\mu}{\epsilon}\right)$ or $\mathcal{O}\left(\log \frac{n\mu}{\epsilon}\right)$ iterations
- Inexact system solves take $\widetilde{\mathcal{O}}\left(m^3 + \text{nnz}(\mathbf{A})\right)$ time (ignoring $\log$ factors)

Recap

Motivation: Predictor-corrector is a theoretically and empirically fast method for linear programming, but previous theory using direct/exact solvers does not scale.

Structural conditions

- We provide conditions on inexactly computing the PC steps so that the outer iteration complex remains $\mathcal{O}\left(\sqrt{n}\log\frac{\mu_0}{\epsilon}\right)$ and the returned solution is ϵ-feasible
- We provide conditions on inexactly computing the PC step and a *correction vector* so that slightly modifying the PC algo. returns an exactly feasible solution while outer iteration complexity remains $\mathcal{O}\left(\sqrt{n}\log\frac{\mu_0}{\epsilon}\right)$.

Efficient iterative solvers

- Construct a strong preconditioner using sketching
- Each iteration of the predictor-corrector method then takes $\widetilde{\mathcal{O}}\left(m^3 + \text{nnz}(\mathbf{A})\right)$ time.

Thank you!

Questions?

References

Agniva Chowdhury, Palma London, Haim Avron, and Petros Drineas. Faster randomized infeasible interior point methods for tall/wide linear programs. Advances in Neural Information Processing Systems, 33:8704–8715, 2020.

Goran Lesaja. Introducing interior-point methods for introductory operations research courses and/or linear programming courses. Open Operational Research Journal, 3:1, 2009.

Stephen J Wright. Primal-dual interior-point methods. SIAM, 1997.