Beyond Convenience: Beyond Convexity

Purushottam Kar

MINI-SYMPOSIUM ON COMPUTATION AND OPTIMIZATION IN THE SCIENCES AND ENGINEERING

Outline of the Talk

- Convex Optimization
- A Few Contemporary Applications
- Non-convex Optimization
- Robust Regression
- Applications of Robust Regression
- Robust PCA

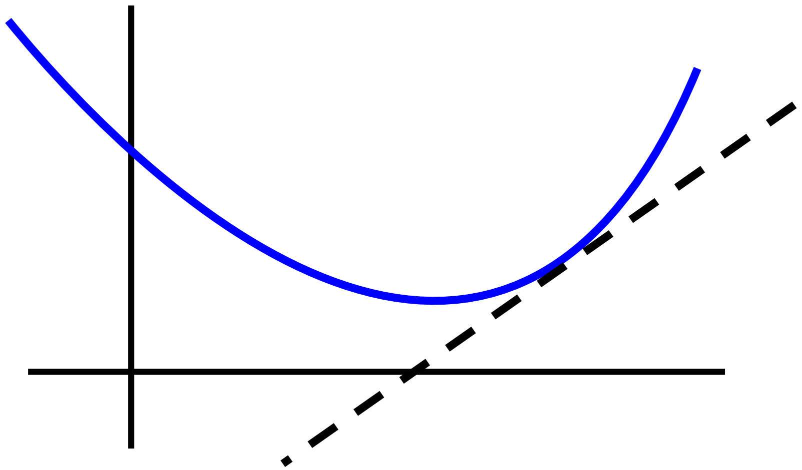
Convex Optimization

Convex Optimization

$$\min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x})$$

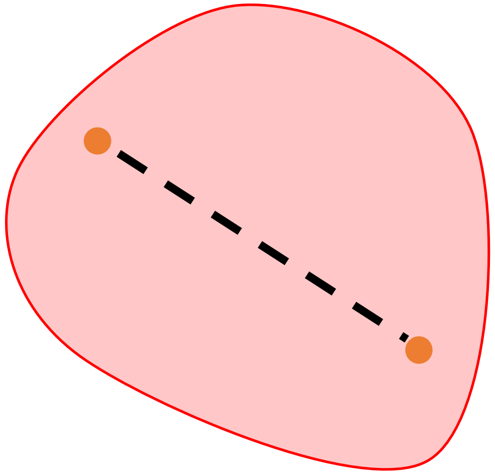
$$f : \mathbb{R}^d \to \mathbb{R}$$

Convex function



$$\mathcal{C} \subseteq \mathbb{R}^d$$

Convex set



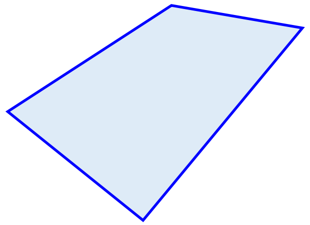
Examples

Linear Programming

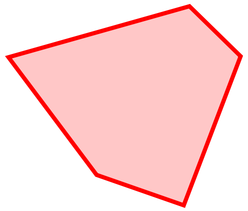
$$\min_{\mathbf{x}\in\mathbb{R}^d} \ \mathbf{a}^\top \mathbf{x}$$

$$s.t. \ \mathbf{b}_i^\top \mathbf{x} \leq c_i$$

f



$\mathcal{C}$

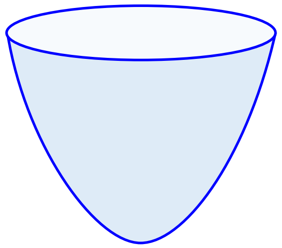


Quadratic Programming

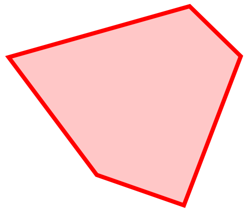
$$\min_{\mathbf{x}\in\mathbb{R}^d} \ \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{a}^\top \mathbf{x}$$

$$s.t. \ \mathbf{b}_i^\top \mathbf{x} \leq c_i$$

f



$\mathcal{C}$

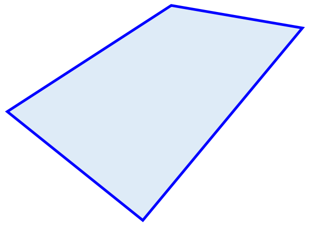


Semidefinite Programming

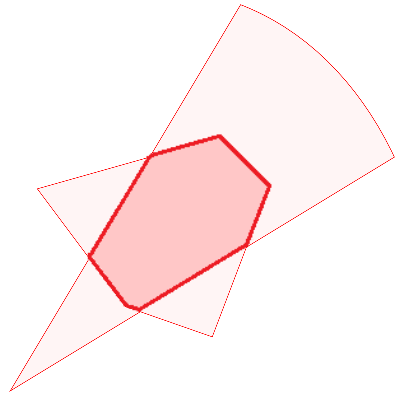
$$\min_{\mathbf{X}\succeq\mathbf{0}} \ \mathbf{A}^\top \mathbf{X}$$

$$s.t. \ \mathbf{B}_i^\top \mathbf{X} \leq c_i$$

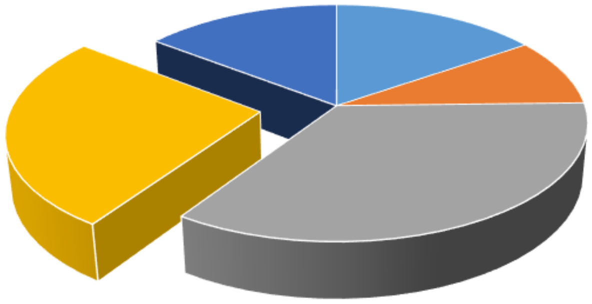
f



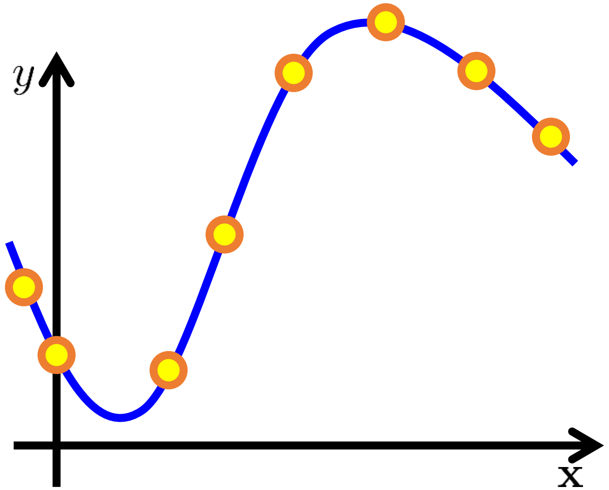
$\mathcal{C}$



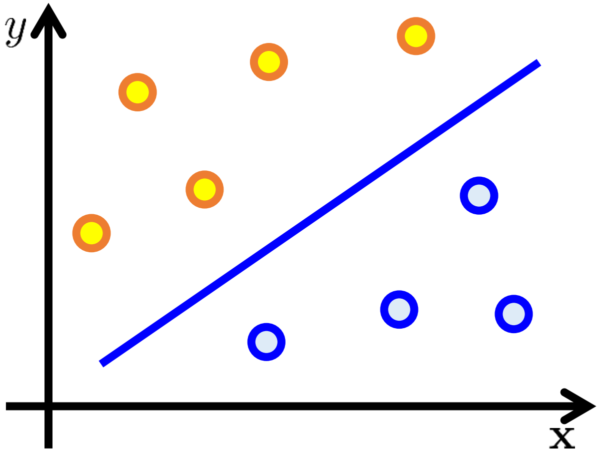
Applications



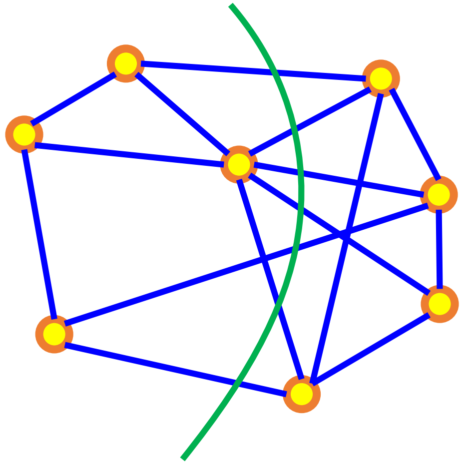
Resource Allocation



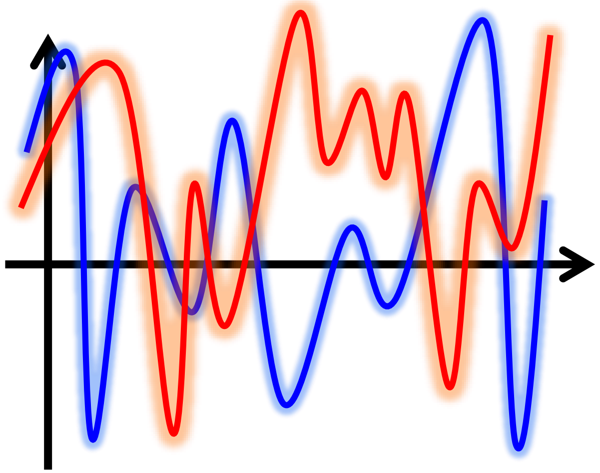
Regression



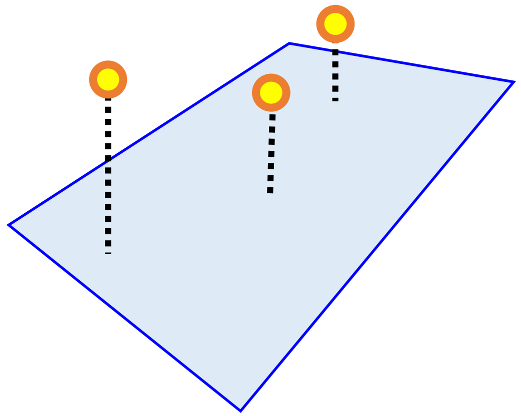
Classification



Clustering/Partitioning



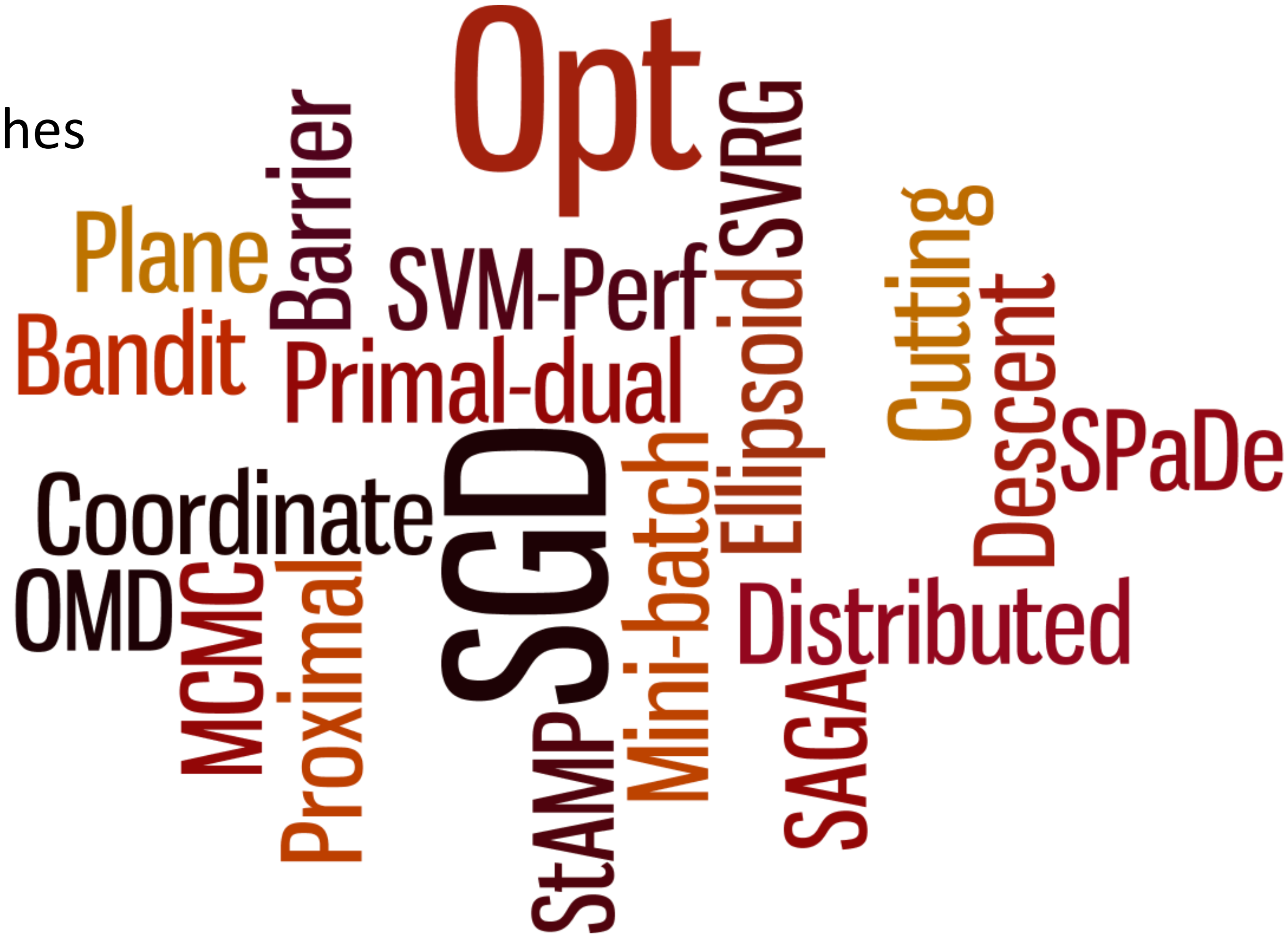
Signal Processing



Dimensionality Reduction

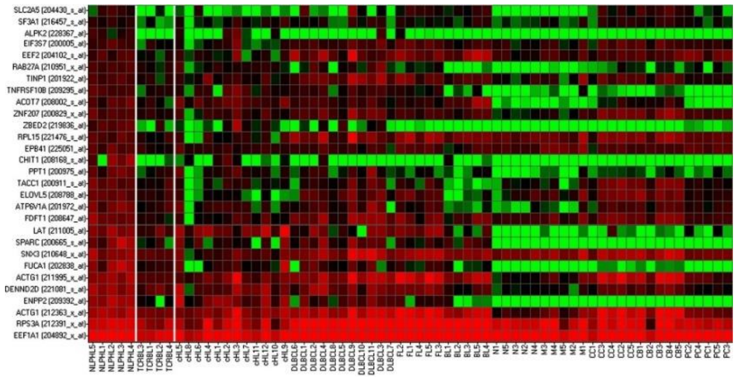
Techniques

- Projected (Sub)gradient Methods
  - Stochastic, mini-batch variants
  - Primal, dual, primal-dual approaches
  - Coordinate update techniques
- Interior Point Methods
  - Barrier methods
  - Annealing methods
- Other Methods
  - Cutting plane methods
  - Accelerated routines
  - Proximal methods
  - Distributed optimization
  - Derivative-free optimization



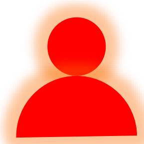
A Few Contemporary Applications

Gene Expression Analysis



DNA micro-array gene expression data

… n subjects



$$_i = (\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$$

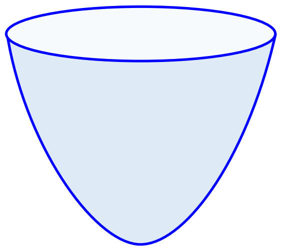
Linear model: $y_i = \mathbf{x}_i^\top \mathbf{w}^*$

Challenge: $\mathbf{w}^*$ is sparse!

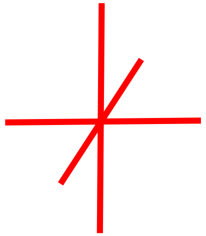
www.tes.com

$$\min_{\mathbf{w} \in \mathcal{B}_0^p(s)} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \mathbf{w})^2$$

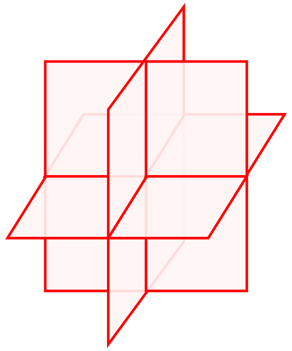
f



$\mathcal{C}$

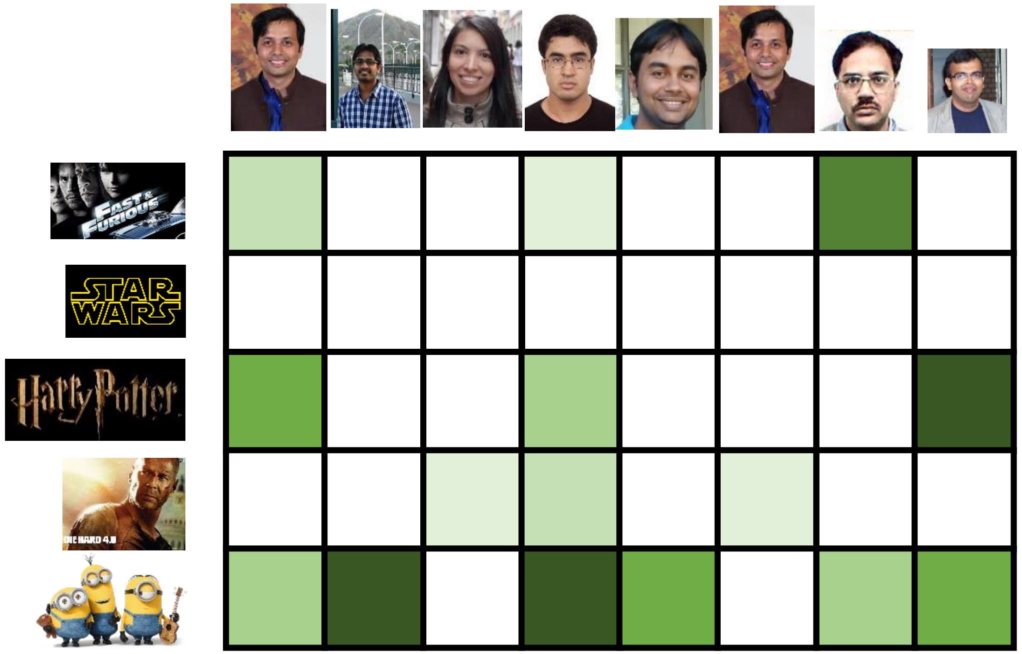


$p = 3, s = 1$

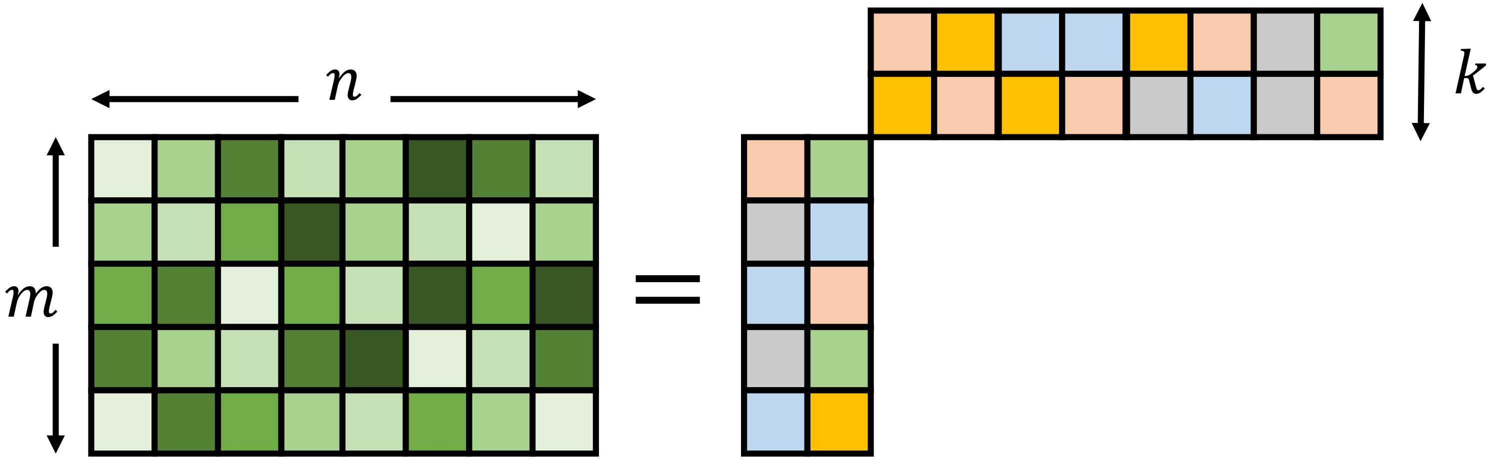


$p = 3, s = 2$

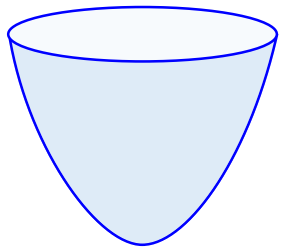
Recommender Systems



$$\min_{L\in\mathcal{M}_k^{m,n}} \|X_\Omega - L_\Omega\|_F^2$$



f



$\mathcal{C}$

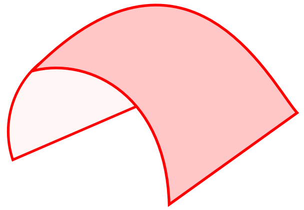


Image Reconstruction and Robust Face Recognition

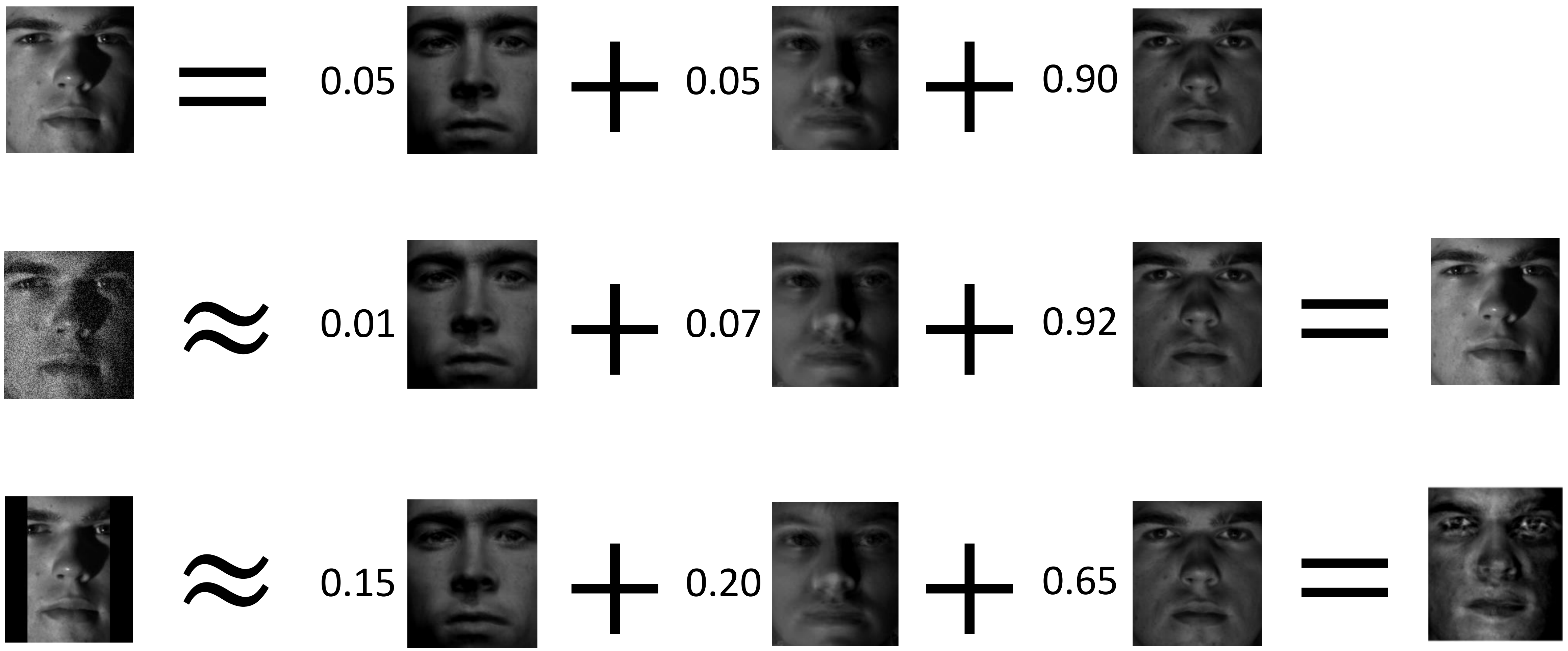
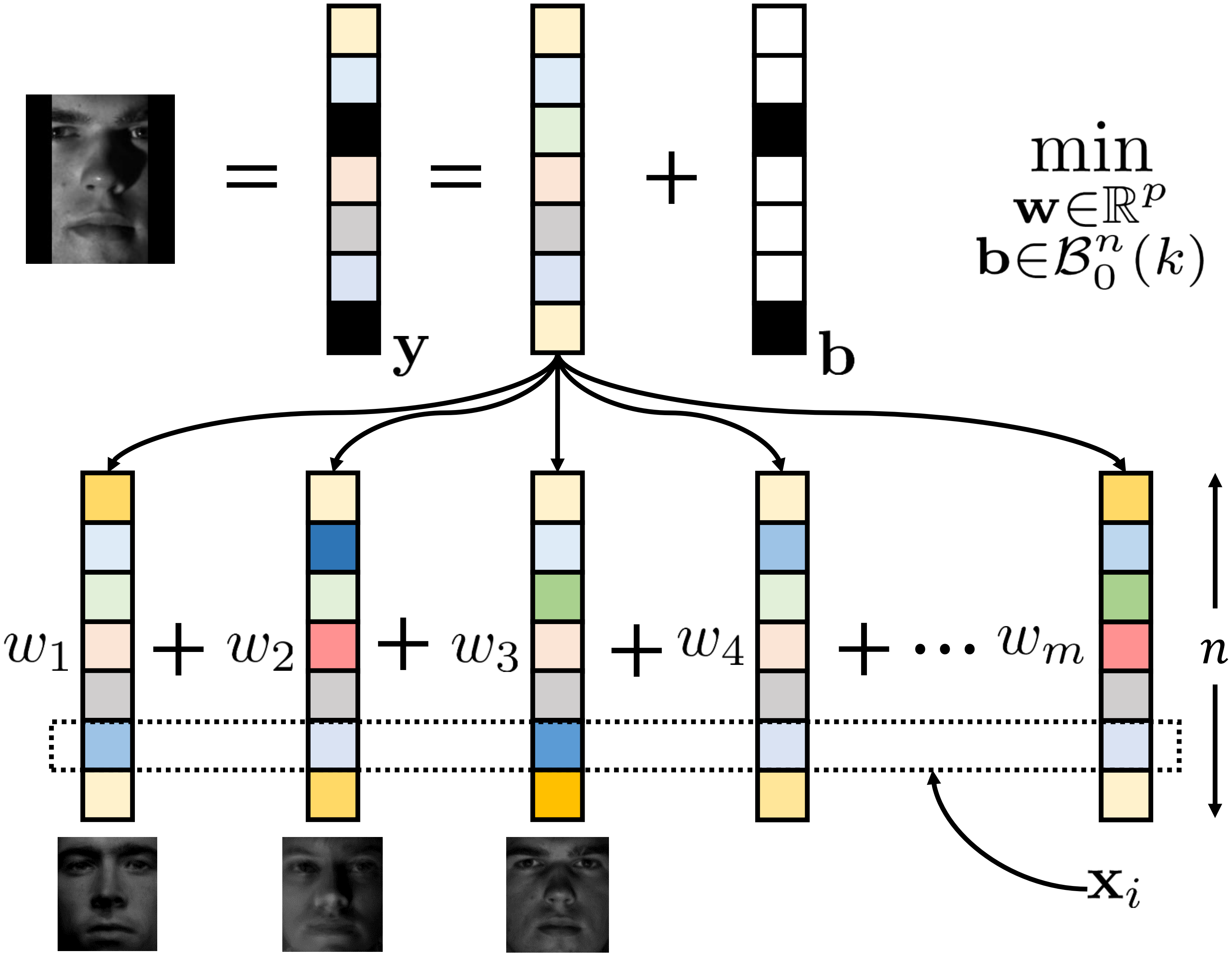
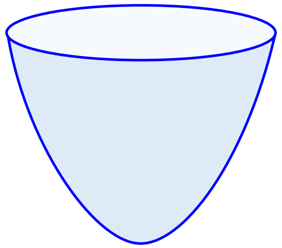


Image Denoising and Robust Face Recognition

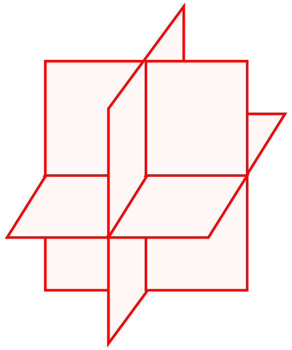


$$\min_{\substack{\mathbf{w}\in\mathbb{R}^p \\ \mathbf{b}\in\mathcal{B}_0^n(k)}} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \mathbf{w} - b_i)^2$$

f

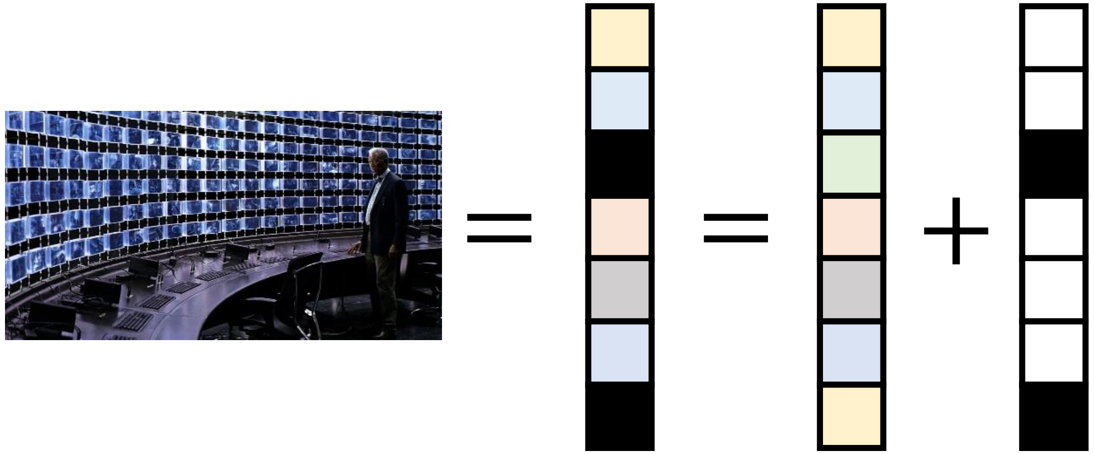


$\mathcal{C}$

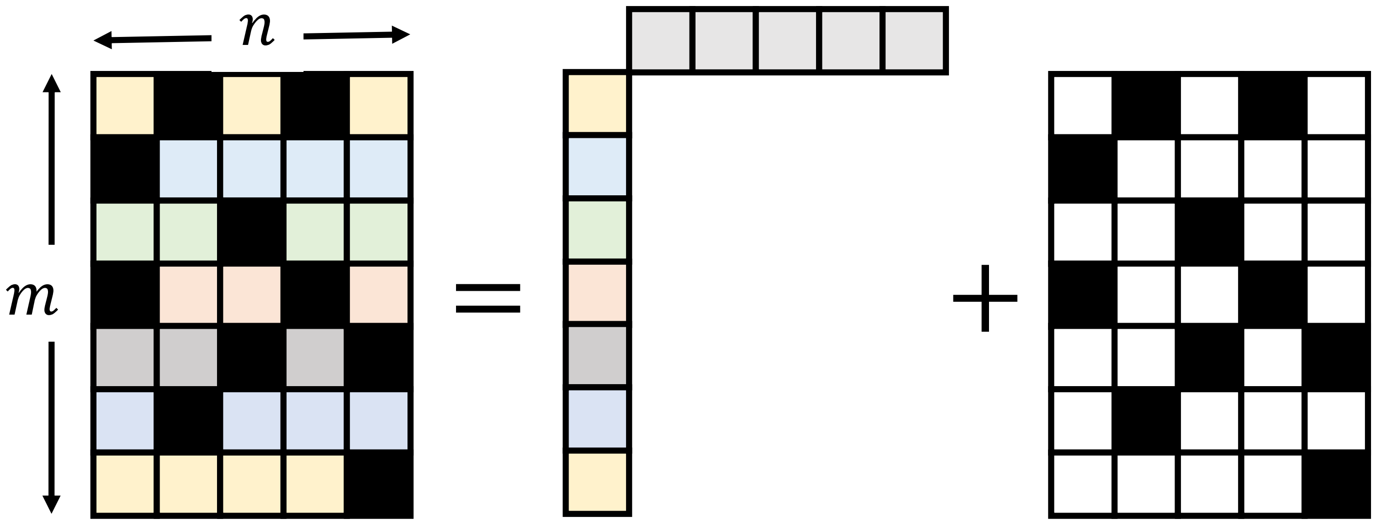


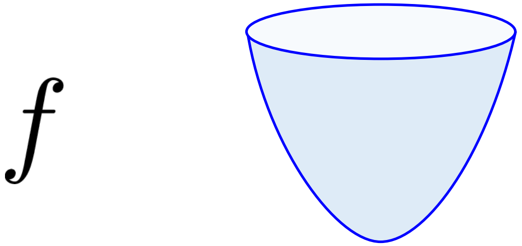
Large Scale Surveillance

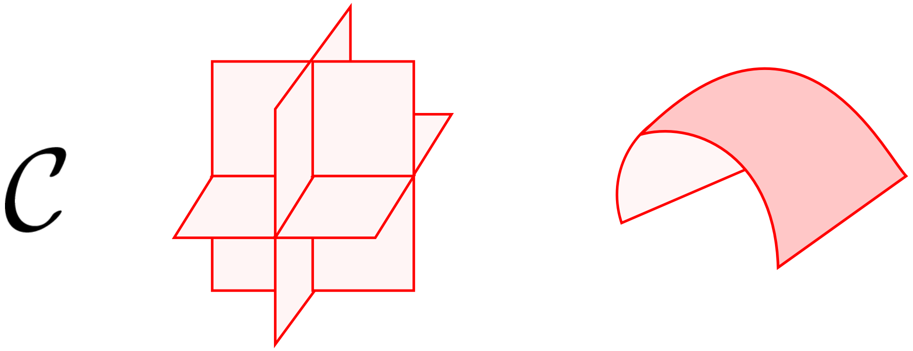
- Foreground-background separation



$$\min_{\substack{L \in \mathcal{M}_k^{m,n} \\ S \in \mathcal{B}_0^{m,n}(s)}} \|X - (L + S)\|_F^2$$



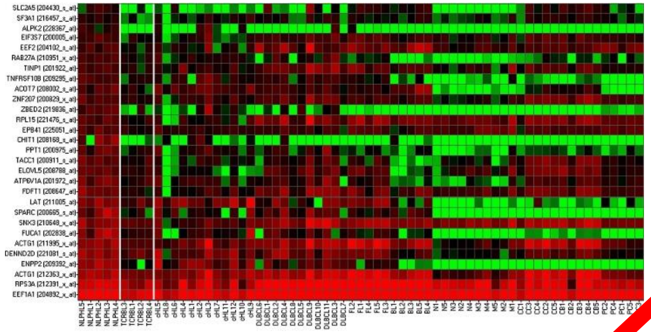




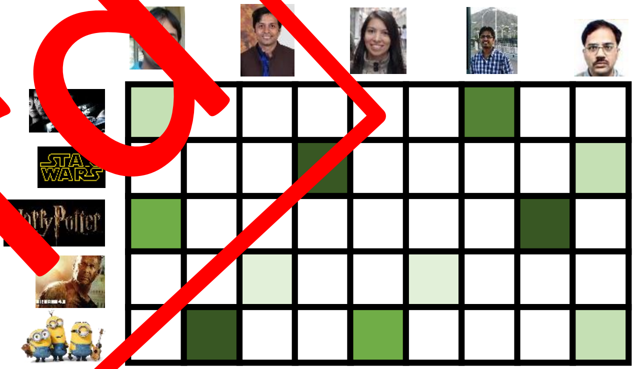
www.extremetech.com

Non Convex Optimization

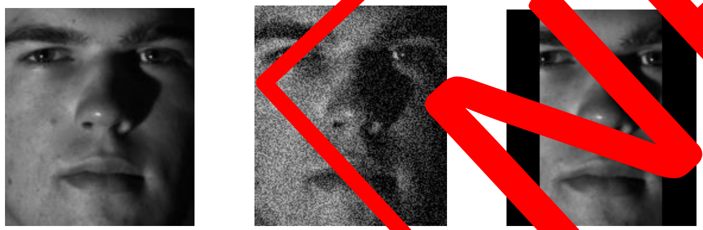
NP-hard



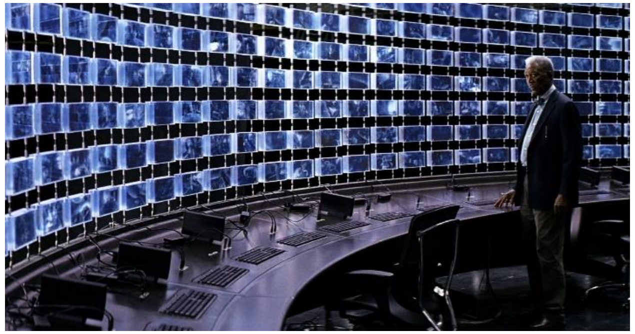
Sparse Recovery



Matrix Completion



Robust Regression

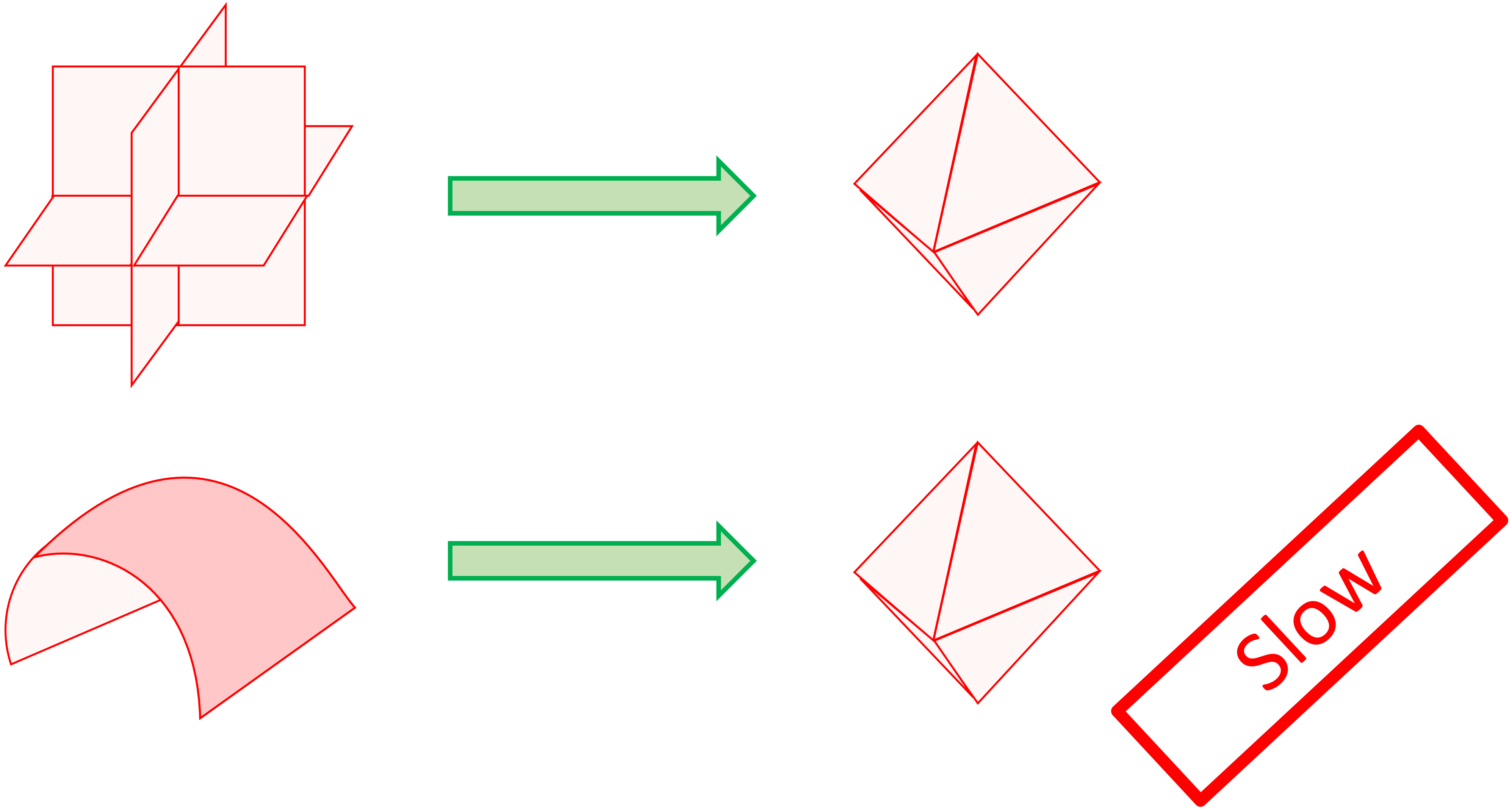


Robust PCA

Non-convex Optimization

Relaxation-based Techniques

- “Convexify” the feasible set



Alternating Minimization

$$\min \; f(\mathbf{x}, \mathbf{y})$$
$$s.t. \; \mathbf{x} \in \mathcal{C}_1$$
$$\mathbf{y} \in \mathcal{C}_2$$

- $\triangleright$ Initialize $\mathbf{x}^0, \mathbf{y}^0$
- $\triangleright$ For $t = 1, 2, \ldots$
  - $\triangleright$ $\mathbf{x}^t = \arg\min_{\mathbf{x} \in \mathcal{C}_1} f(\mathbf{x}, \mathbf{y}^{t-1})$
  - $\triangleright$ $\mathbf{y}^t = \arg\min_{\mathbf{y} \in \mathcal{C}_2} f(\mathbf{x}^t, \mathbf{y})$

Matrix Completion

$$\min_{L \in \mathcal{M}_k^{m,n}} \|X_\Omega - L_\Omega\|_F^2$$
$$\equiv \min_{\substack{U \in \mathbb{R}^{m \times k} \\ V \in \mathbb{R}^{n \times k}}} \|X_\Omega - (UV^\top)_\Omega\|_F^2$$

Robust PCA

$$\min_{\substack{L \in \mathcal{M}_k^{m,n} \\ S \in \mathcal{B}_0^{m,n}(s)}} \|X - (L + S)\|_F^2$$

… also Robust Regression, coming up

Projected Gradient Descent

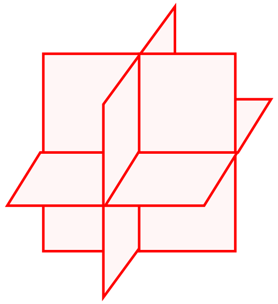
$$\min \; f(\mathbf{x})$$
$$s.t. \; \mathbf{x} \in \mathcal{C}$$

- $\triangleright$ Initialize $\mathbf{x}^0$
- $\triangleright$ For $t = 1, 2, \ldots$
  - $\triangleright$ $\mathbf{z}^t = \mathbf{x}^{t-1} - \eta_t \cdot \nabla f(\mathbf{x}^{t-1})$
  - $\triangleright$ $\mathbf{x}^t = \Pi_{\mathcal{C}}(\mathbf{z}^t)$

$$\Pi_{\mathcal{C}}(\mathbf{z}) = \arg\min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{z} - \mathbf{x}\|_2^2$$

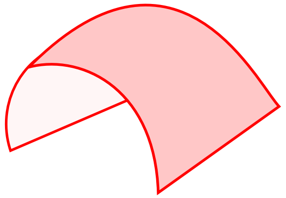
Non-convex Projection

$\mathcal{B}_0^p(s)$



Top s elements by magnitude

$\mathcal{M}_k^{m,n}$



Perform k-truncated SVD

Sparse Recovery

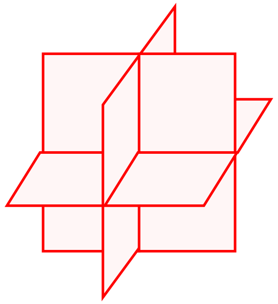
$$\min_{\mathbf{w} \in \mathcal{B}_0^p(s)} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \mathbf{w})^2$$

Pursuit and Greedy Methods

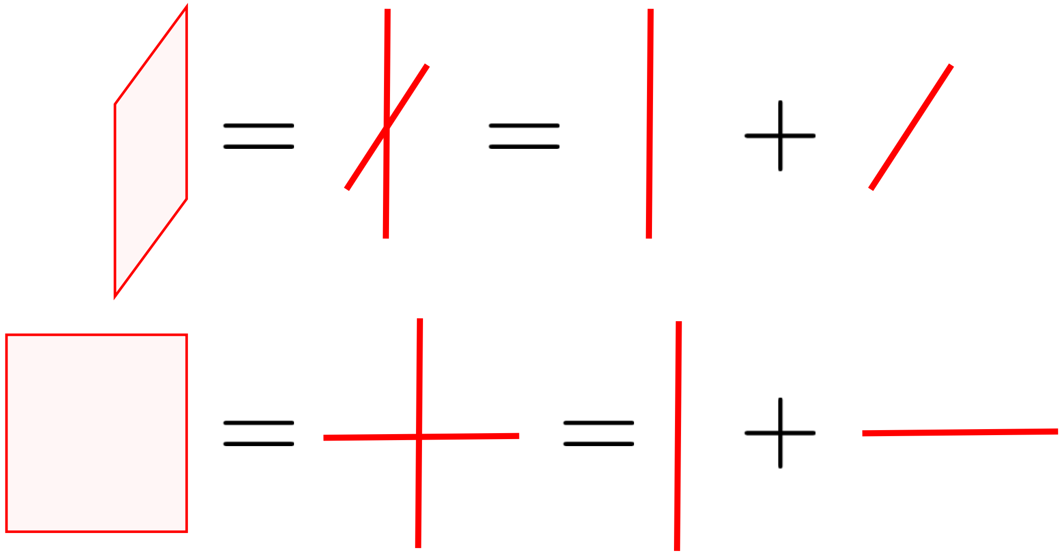
$$\min \; f(\mathbf{x})$$
$$s.t. \; \mathbf{x} \in \mathcal{C}$$

$\mathcal{A}$ Set of “atoms”

$$\mathcal{C} = \left\{ \mathbf{x} = \sum_{i=1}^{s} \mathbf{a}_i : \mathbf{a}_i \in \mathcal{A} \right\}$$



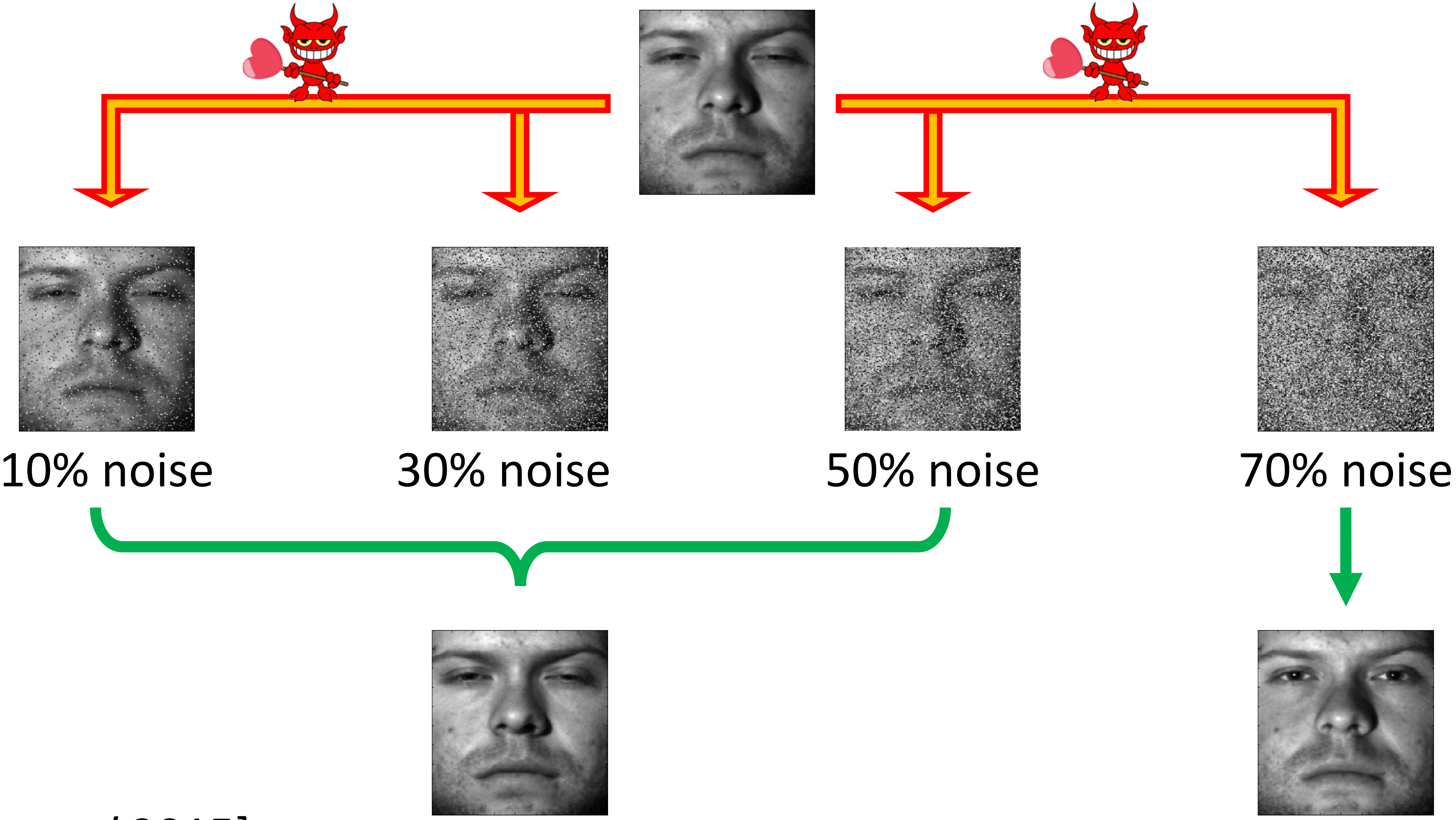
Sparse Recovery



- ▷ Initialize $S^0 = \phi$
- ▷ For $t = 1, 2, \ldots$
  - ▷ $\mathbf{a}^t =$ “best” greedy choice
  - ▷ $S^t = S^{t-1} \cup \{\mathbf{a}^t\}$
  - ▷ $\mathbf{x}^t = \underset{\mathbf{x} \in \text{conv}(S^t)}{\arg\min} \; f(\mathbf{x})$

Applications of NCOpt

Face Recognition



[Bhatia *et al* 2015]

Image Reconstruction

Original | Input | Ordinary LS | Alt-Min

[Bhatia *et al* 2015]

Foreground-background Separation

Convex Relaxation. Runtime: 1700 sec

Alt-Proj. Runtime: 70 sec

[Netrapalli *et al* 2014]

Concluding Comments

Non-convex optimization is an exciting area

Widespread applications

- Much better modelling of problems
- Much more scalable algorithms
- Provable guarantees

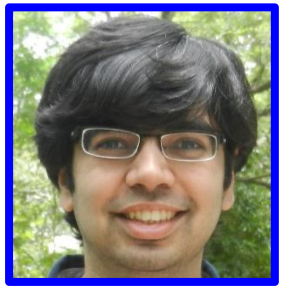
So …

- Full of opportunities
- Full of challenges

Acknowledgements

http://research.microsoft.com/en-us/projects/altmin/default.aspx

Portions of this talk were based on joint work with



Kush Bhatia
Microsoft Research

Prateek Jain
Microsoft Research

Ambuj Tewari
U. Michigan, Ann Arbor

Questions?