Spectral Frank-Wolfe Algorithm: Strict Complementarity and Linear Convergence

Lijun Ding

Joint work with Yingjie Fei, Qiantong Xu, and Chengrun Yang

June 15, 2020

Overview

Convex smooth minimization over a spectrahedron

Main optimization problem:

Convex smooth minimization over a spectrahedron

Main optimization problem:

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- function g strongly convex and smooth

Convex smooth minimization over a spectrahedron

Main optimization problem:

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- function g strongly convex and smooth
- linear map $\mathcal{A}$ and matrix $C \in \mathbf{S}^n$

Convex smooth minimization over a spectrahedron

Main optimization problem:

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- function g strongly convex and smooth
- linear map $\mathcal{A}$ and matrix $C \in \mathbf{S}^n$
- trace $\mathbf{tr}(\cdot)$, sum of diagonals

Convex smooth minimization over a spectrahedron

Main optimization problem:

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- function g strongly convex and smooth
- linear map $\mathcal{A}$ and matrix $C \in \mathbf{S}^n$
- trace $\mathbf{tr}(\cdot)$, sum of diagonals
- positive semidefinite matrices $\mathbf{S}^n_+$, i.e., symmetric matrices with non-negative eigenvalues

Convex smooth minimization over a spectrahedron

Main optimization problem:

$$\begin{array}{ll} \underset{X\in\mathbf{S}^n\subset\mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- function g strongly convex and smooth
- linear map $\mathcal{A}$ and matrix $C \in \mathbf{S}^n$
- trace $\mathbf{tr}(\cdot)$, sum of diagonals
- positive semidefinite matrices $\mathbf{S}^n_+$, i.e., symmetric matrices with non-negative eigenvalues
- **unique** optimal solution $X_\star$

Applications

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

Applications

$$\begin{array}{ll} \underset{X\in \mathbf{S}^n\subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \qquad \text{(M)}$$

- matrix sensing [RFP10]

Applications

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- matrix sensing [RFP10]
- matrix completion [CR09, JS10]

Applications

$$\begin{array}{ll} \underset{X\in\mathbf{S}^n\subset\mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- matrix sensing [RFP10]
- matrix completion [CR09, JS10]
- phase retrieval [CESV15, YUTC17]

Applications

$$\begin{array}{ll} \underset{X\in\mathbf{S}^n\subset\mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- matrix sensing [RFP10]
- matrix completion [CR09, JS10]
- phase retrieval [CESV15, YUTC17]
- one-bit matrix completion [DPVDBW14]

Applications

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n \times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- matrix sensing [RFP10]
- matrix completion [CR09, JS10]
- phase retrieval [CESV15, YUTC17]
- one-bit matrix completion [DPVDBW14]
- blind deconvolution [ARR13]

Applications

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n \subset \mathbf{R}^{n\times n}}{\text{minimize}} & f(X) := g(\mathcal{A}X) + \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(X) = 1, \quad \text{and} \quad X \in \mathbf{S}^n_+, \end{array} \tag{M}$$

- matrix sensing [RFP10]
- matrix completion [CR09, JS10]
- phase retrieval [CESV15, YUTC17]
- one-bit matrix completion [DPVDBW14]
- blind deconvolution [ARR13]

Expect rank $r_\star = \mathbf{rank}(X_\star) \ll n$!

Projected Gradient (PG)

$$\text{minimize}_{X \in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \tag{M}$$

Projected Gradient (PG)

$$\text{minimize}_{X\in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\, X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \qquad \text{(M)}$$

- orthogonal projection: $\mathcal{P}_{\mathcal{SP}^n}(X) = \arg\min_V \|X - V\|_{\mathrm{F}}$

Projected Gradient (PG)

$$\text{minimize}_{X \in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \tag{M}$$

- orthogonal projection: $\mathcal{P}_{\mathcal{SP}^n}(X) = \arg\min_V \|X - V\|_F$
- PG: Choose $X_0 \in \mathcal{SP}^n$ and $\eta > 0$, iterate

$$X_{t+1} = \mathcal{P}_{\mathcal{SP}^n}\left(X_t - \eta \nabla f(X_t)\right). \tag{PG}$$

Projected Gradient (PG)

$$\text{minimize}_{X \in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \tag{M}$$

- orthogonal projection: $\mathcal{P}_{\mathcal{SP}^n}(X) = \arg\min_V \|X - V\|_F$
- PG: Choose $X_0 \in \mathcal{SP}^n$ and $\eta > 0$, iterate

$$X_{t+1} = \mathcal{P}_{\mathcal{SP}^n}\left(X_t - \eta \nabla f(X_t)\right). \tag{PG}$$

- iteration complexity $\mathcal{O}(\frac{1}{\epsilon})$

Projected Gradient (PG)

$$\text{minimize}_{X \in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \tag{M}$$

- orthogonal projection: $\mathcal{P}_{\mathcal{SP}^n}(X) = \arg\min_V \|X - V\|_F$
- PG: Choose $X_0 \in \mathcal{SP}^n$ and $\eta > 0$, iterate

$$X_{t+1} = \mathcal{P}_{\mathcal{SP}^n}\left(X_t - \eta \nabla f(X_t)\right). \tag{PG}$$

- iteration complexity $\mathcal{O}(\frac{1}{\epsilon})$
- accelerated PG, $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$

Projected Gradient (PG)

$$\text{minimize}_{X \in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+,}_{\mathcal{SP}^n} \tag{M}$$

- orthogonal projection: $\mathcal{P}_{\mathcal{SP}^n}(X) = \arg\min_V \|X - V\|_{\mathrm{F}}$
- PG: Choose $X_0 \in \mathcal{SP}^n$ and $\eta > 0$, iterate

$$X_{t+1} = \mathcal{P}_{\mathcal{SP}^n}\left(X_t - \eta \nabla f(X_t)\right). \tag{PG}$$

- iteration complexity $\mathcal{O}(\frac{1}{\epsilon})$
- accelerated PG, $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$

Bottleneck: $\mathcal{O}(n^3)$ per iteration due to FULL EVD in $\mathcal{P}_{\mathcal{SP}^n}$!

Projection free method: Frank-Wolfe (FW)

$$\text{minimize}_{X \in \mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \qquad \text{(M)}$$

Projection free method: Frank-Wolfe (FW)

$$\text{minimize}_{X\in\mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X)=1,\, X\in\mathbf{S}^n_+}_{\mathcal{SP}^n}, \qquad \text{(M)}$$

- FW: choose $X_0 \in \mathcal{SP}^n$, iterate

(LOO) Linear Optimization Oracle: $V_t = \arg\min_{V\in\mathcal{SP}^n} \mathbf{tr}(V\nabla f(X_t))$.

Projection free method: Frank-Wolfe (FW)

$$\text{minimize}_{X\in\mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X)=1,\, X\in\mathbf{S}^n_+}_{\mathcal{SP}^n}, \qquad \text{(M)}$$

- FW: choose $X_0 \in \mathcal{SP}^n$, iterate

(LOO) Linear Optimization Oracle: $V_t = \arg\min_{V\in\mathcal{SP}^n} \mathbf{tr}(V\nabla f(X_t))$.
(LS) Line Search: X_{t+1} solves $\min_{X=\eta X_t+(1-\eta)V_t,\eta\in[0,1]} f(X)$.

- **Low per iteration complexity**: LOO only needs to compute one eigenvector of $\nabla f(X_t)$!

Projection free method: Frank-Wolfe (FW)

$$\text{minimize}_{X\in\mathbf{S}^n} \quad f(X) \quad \text{subject to} \quad \underbrace{\mathbf{tr}(X) = 1,\ X \in \mathbf{S}^n_+}_{\mathcal{SP}^n}, \qquad \text{(M)}$$

- FW: choose $X_0 \in \mathcal{SP}^n$, iterate

(LOO) Linear Optimization Oracle: $V_t = \arg\min_{V\in\mathcal{SP}^n} \mathbf{tr}(V\nabla f(X_t))$.
(LS) Line Search: X_{t+1} solves $\min_{X=\eta X_t+(1-\eta)V_t,\eta\in[0,1]} f(X)$.

- **Low per iteration complexity**: LOO only needs to compute one eigenvector of $\nabla f(X_t)$!

Bottleneck: Slow convergence, $\mathcal{O}(\frac{1}{\epsilon})$ iteration complexity in both theory and practice!

FW variants

Many variants:

- Randomized regularized FW [Gar16]
- In-face direction FW [FGM17]
- BlockFW [AZHHL17]
- FW with $r_\star = \textbf{rank}(X_\star) = 1$ [Gar19]

Shortage: No linear convergence or sensitive to input rank estimate or $r_\star = 1$.

Outline

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \ldots, v_k] \in \mathbf{R}^{n \times k}$ of $\nabla f(X_t)$.

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \dots, v_k] \in \mathbf{R}^{n\times k}$ of $\nabla f(X_t)$.
- k Spectral Search (kSS): $X_{t+1} = \eta_\star X_t + VS_\star V^\top$, in which $\eta_\star \in \mathbf{R}, S_\star \in \mathbf{S}^k$ solves

$$\min f(\eta X_t + VSV^\top) \quad \text{s.t.} \quad S \in \mathbf{S}_+^k, \eta + \mathbf{tr}(S) = 1, \eta \geq 0.$$

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \ldots, v_k] \in \mathbf{R}^{n\times k}$ of $\nabla f(X_t)$.
- k Spectral Search (kSS): $X_{t+1} = \eta_\star X_t + V S_\star V^\top$, in which $\eta_\star \in \mathbf{R}, S_\star \in \mathbf{S}^k$ solves

$$\min f(\eta X_t + VSV^\top) \quad \text{s.t.} \quad S \in \mathbf{S}^k_+, \eta + \mathbf{tr}(S) = 1, \eta \geq 0.$$

Both procedure are easy to solve for small k!

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \dots, v_k] \in \mathbf{R}^{n \times k}$ of $\nabla f(X_t)$.
- k Spectral Search (kSS): $X_{t+1} = \eta_\star X_t + V S_\star V^\top$, in which $\eta_\star \in \mathbf{R}, S_\star \in \mathbf{S}^k$ solves

$$\min f(\eta X_t + VSV^\top) \quad \text{s.t.} \quad S \in \mathbf{S}_+^k, \eta + \mathbf{tr}(S) = 1, \eta \geq 0.$$

Both procedure are easy to solve for small k!

Moreover...

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \ldots, v_k] \in \mathbf{R}^{n \times k}$ of $\nabla f(X_t)$.
- k Spectral Search (kSS): $X_{t+1} = \eta_\star X_t + V S_\star V^\top$, in which $\eta_\star \in \mathbf{R}, S_\star \in \mathbf{S}^k$ solves

$$\min f(\eta X_t + VSV^\top) \quad \text{s.t.} \quad S \in \mathbf{S}_+^k, \eta + \mathbf{tr}(S) = 1, \eta \geq 0.$$

Both procedure are easy to solve for small k!

Moreover...

- $\mathcal{O}(\frac{1}{\epsilon})$ convergence for general k.

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \ldots, v_k] \in \mathbf{R}^{n \times k}$ of $\nabla f(X_t)$.
- k Spectral Search (kSS): $X_{t+1} = \eta_\star X_t + V S_\star V^\top$, in which $\eta_\star \in \mathbf{R}, S_\star \in \mathbf{S}^k$ solves

$$\min f(\eta X_t + VSV^\top) \quad \text{s.t.} \quad S \in \mathbf{S}^k_+, \eta + \mathbf{tr}(S) = 1, \eta \geq 0.$$

Both procedure are easy to solve for small k!

Moreover...

- $\mathcal{O}(\frac{1}{\epsilon})$ convergence for general k.
- **Linear convergence if $k \geq r_\star$!**

Spectral Frank-Wolfe (SpecFW)

Spectral Frank-Wolfe: choose $X_0 \in \mathcal{SP}^n$, a rank estimate $k > 0$, iterate

- kLOO: Compute bottom k eigenvectors $V = [v_1, \dots, v_k] \in \mathbf{R}^{n\times k}$ of $\nabla f(X_t)$.
- k Spectral Search (kSS): $X_{t+1} = \eta_\star X_t + VS_\star V^\top$, in which $\eta_\star \in \mathbf{R}$, $S_\star \in \mathbf{S}^k$ solves

$$\min f(\eta X_t + VSV^\top) \quad \text{s.t.} \quad S \in \mathbf{S}_+^k, \eta + \mathbf{tr}(S) = 1, \eta \geq 0.$$

Both procedure are easy to solve for small k!

Moreover...

- $\mathcal{O}(\frac{1}{\epsilon})$ convergence for general k.
- **Linear convergence if $k \geq r_\star$!** (also needs strict complementarity)

Comparison with FW

Two stronger subproblem oracles:

Table: Comparison with FW

FW	**SpecFW**
LOO: Compute one eigenvector v	kLOO: Compute k eigenvectors V

Comparison with FW

Two stronger subproblem oracles:

Table: Comparison with FW

FW	SpecFW
LOO: Compute one eigenvector v	kLOO: Compute k eigenvectors V
Line Search (LS): $\min f(\eta X_t + (1-\eta)vv^\top)$ s.t. $\eta \in [0,1]$	k Spectral Search (kSS): $\min f(\eta X_t + VSV^\top)$ s.t. $\eta \geq 0, S \in \mathbf{S}_+^k, \mathbf{tr}(S) + \eta = 1$

Comparison with FW

Two stronger subproblem oracles:

Table: Comparison with FW

FW	**SpecFW**
LOO: Compute one eigenvector v	kLOO: Compute k eigenvectors V
Line Search (LS): $\min f(\eta X_t + (1-\eta)vv^\top)$ s.t. $\eta \in [0,1]$	k Spectral Search (kSS): $\min f(\eta X_t + VSV^\top)$ s.t. $\eta \geq 0, S \in \mathbf{S}_+^k, \mathbf{tr}(S) + \eta = 1$

In fact, when $k = 1$, SpecFW *is* FW!

Comparison with FW

Two stronger subproblem oracles:

Table: Comparison with FW

FW	**SpecFW**
LOO: Compute one eigenvector v	kLOO: Compute k eigenvectors V
Line Search (LS): $\min f(\eta X_t + (1-\eta)vv^\top)$ s.t. $\eta \in [0,1]$	k Spectral Search (kSS): $\min f(\eta X_t + VSV^\top)$ s.t. $\eta \geq 0, S \in \mathbf{S}_+^k, \mathbf{tr}(S) + \eta = 1$

In fact, when $k = 1$, SpecFW *is* FW! Expect *at least* $\mathcal{O}(\frac{1}{\epsilon})$ convergence even if $k \leq r_\star$.

Comparison with FW

Two stronger subproblem oracles:

Table: Comparison with FW

FW	SpecFW
LOO: Compute one eigenvector v	kLOO: Compute k eigenvectors V
Line Search (LS): $\min f(\eta X_t + (1-\eta)vv^\top)$ s.t. $\eta \in [0,1]$	k Spectral Search (kSS): $\min f(\eta X_t + VSV^\top)$ s.t. $\eta \geq 0, S \in \mathbf{S}_+^k, \mathbf{tr}(S) + \eta = 1$

In fact, when $k = 1$, SpecFW *is* FW! Expect *at least* $\mathcal{O}(\frac{1}{\epsilon})$ convergence even if $k \leq r_\star$.

How about linear convergence when $k \geq r_\star$?

Two stronger subproblem oracles:

Table: Comparison with FW

FW	**SpecFW**
LOO: Compute one eigenvector v	kLOO: Compute k eigenvectors V
Line Search (LS): $\min f(\eta X_t + (1-\eta)vv^\top)$ s.t. $\eta \in [0,1]$	k Spectral Search (kSS): $\min f(\eta X_t + VSV^\top)$ s.t. $\eta \geq 0, S \in \mathbf{S}_+^k, \mathbf{tr}(S) + \eta = 1$

In fact, when $k = 1$, SpecFW *is* FW! Expect *at least* $\mathcal{O}(\frac{1}{\epsilon})$ convergence even if $k \leq r_\star$.

How about linear convergence when $k \geq r_\star$?
What is strict complementarity?

Strict complementarity

Eigenspace of $\nabla f(X_\star)$ for the smallest eigenvalue, $\mathbf{EV}(\nabla f(X_\star)) \subset \mathbf{R}^n$

Strict complementarity

Eigenspace of $\nabla f(X_\star)$ for the smallest eigenvalue, $\mathbf{EV}(\nabla f(X_\star)) \subset \mathbf{R}^n$

$$\text{KKT} \implies \mathbf{range}(X_\star) \subset \mathbf{EV}(\nabla f(X_\star))$$

Strict complementarity

Eigenspace of $\nabla f(X_\star)$ for the smallest eigenvalue, $\mathbf{EV}(\nabla f(X_\star)) \subset \mathbf{R}^n$

$$\begin{aligned} \text{KKT} &\implies \textbf{range}(X_\star) \subset \mathbf{EV}(\nabla f(X_\star)) \\ &\implies \underbrace{\dim(\textbf{range}(X_\star))}_{=:r_\star} \leq \underbrace{\dim(\mathbf{EV}(\nabla f(X_\star)))}_{=:k_\star} \end{aligned}$$

Note that the smallest eigenvalue has multiplicity at least $r_\star$:

$$\lambda_{n-r_\star+1}(\nabla f(X_\star)) = \cdots = \lambda_n(\nabla f(X_\star)).$$

Here $\lambda_{n-i+1}(\nabla f(X_\star))$ is the i-th smallest eigenvalue.

Strict complementarity

Eigenspace of $\nabla f(X_\star)$ for the smallest eigenvalue, $\mathbf{EV}(\nabla f(X_\star)) \subset \mathbf{R}^n$

$$\begin{aligned} \text{KKT} &\implies \mathbf{range}(X_\star) \subset \mathbf{EV}(\nabla f(X_\star)) \\ &\implies \underbrace{\dim(\mathbf{range}(X_\star))}_{=:r_\star} \leq \underbrace{\dim(\mathbf{EV}(\nabla f(X_\star)))}_{=:k_\star} \end{aligned}$$

Note that the smallest eigenvalue has multiplicity at least $r_\star$:

$$\lambda_{n-r_\star+1}(\nabla f(X_\star)) = \cdots = \lambda_n(\nabla f(X_\star)).$$

Here $\lambda_{n-i+1}(\nabla f(X_\star))$ is the i-th smallest eigenvalue.

Strict complementarity (st. comp.) is $r_\star = k_\star$.

Strict complementarity

Eigenspace of $\nabla f(X_\star)$ for the smallest eigenvalue, $\mathbf{EV}(\nabla f(X_\star)) \subset \mathbf{R}^n$

$$
\begin{aligned}
\text{KKT} &\implies \mathbf{range}(X_\star) \subset \mathbf{EV}(\nabla f(X_\star)) \\
&\implies \underbrace{\dim(\mathbf{range}(X_\star))}_{=:r_\star} \leq \underbrace{\dim(\mathbf{EV}(\nabla f(X_\star)))}_{=:k_\star}
\end{aligned}
$$

Note that the smallest eigenvalue has multiplicity at least $r_\star$:

$$\lambda_{n-r_\star+1}(\nabla f(X_\star)) = \cdots = \lambda_n(\nabla f(X_\star)).$$

Here $\lambda_{n-i+1}(\nabla f(X_\star))$ is the i-th smallest eigenvalue.

Strict complementarity (st. comp.) is $r_\star = k_\star$.

More concretely, st. comp. is an eigengap condition on $r_\star$-th and $r_\star + 1$-th smallest eigenvalue:

$$\lambda_{n-r_\star}(\nabla f(X_\star)) - \lambda_{n-r_\star+1}(\nabla f(X_\star)) > 0.$$

Intuition of linear convergence

Under strict complementarity $r_\star = k_\star$:

Intuition of linear convergence

Under strict complementarity $r_\star = k_\star$:

1. **range**$(X_\star) = \mathbf{EV}(\nabla f(X_\star))$

Intuition of linear convergence

Under strict complementarity $r_\star = k_\star$:

1. $\textbf{range}(X_\star) = \textbf{EV}(\nabla f(X_\star))$
2. Compute $V_\star = [v_1, \ldots, v_{k_\star}]$, the bottom eigenvectors of $\nabla f(X_\star)$.

Intuition of linear convergence

Under strict complementarity $r_\star = k_\star$:

1. $\textbf{range}(X_\star) = \textbf{EV}(\nabla f(X_\star))$
2. Compute $V_\star = [v_1, \ldots, v_{k_\star}]$, the bottom eigenvectors of $\nabla f(X_\star)$.
3. $X_\star = V_\star S_\star V_\star^\top$ for some $S_\star \in \mathbf{S}_+^{r_\star}$, $\textbf{tr}(S) = 1$

Intuition of linear convergence

Under strict complementarity $r_\star = k_\star$:

1. $\textbf{range}(X_\star) = \textbf{EV}(\nabla f(X_\star))$
2. Compute $V_\star = [v_1, \ldots, v_{k_\star}]$, the bottom eigenvectors of $\nabla f(X_\star)$.
3. $X_\star = V_\star S_\star V_\star^\top$ for some $S_\star \in \mathbf{S}_+^{r_\star}$, $\textbf{tr}(S) = 1$
4. Obtain $S_\star$ by solving

$$\text{minimize} \quad f(V_\star S_\star V_\star^\top) \quad \text{s.t.} \quad S \in \mathbf{S}_+^{r_\star}, \textbf{tr}(S) = 1. \qquad \text{(reduced M)}$$

5. Problem (M) is solved given $\nabla f(X_\star)$!

Intuition of linear convergence

Under strict complementarity $r_\star = k_\star$:

1. $\mathbf{range}(X_\star) = \mathbf{EV}(\nabla f(X_\star))$
2. Compute $V_\star = [v_1, \dots, v_{k_\star}]$, the bottom eigenvectors of $\nabla f(X_\star)$.
3. $X_\star = V_\star S_\star V_\star^\top$ for some $S_\star \in \mathbf{S}_+^{r_\star}$, $\mathbf{tr}(S) = 1$
4. Obtain $S_\star$ by solving

$$\text{minimize} \quad f(V_\star S_\star V_\star^\top) \quad \text{s.t.} \quad S \in \mathbf{S}_+^{r_\star}, \mathbf{tr}(S) = 1. \qquad \text{(reduced M)}$$

5. Problem (M) is solved given $\nabla f(X_\star)$!

SpecFW is simply algorithimic procedures for step 2 and 4!

Outline

1. Introduction
   - Problem setup
   - Past algorithms
2. SpecFW and strict complementarity
   - Spectral Frank-Wolfe (SpecFW)
   - Strict complementarity
3. Numerics
   - Experimental setup
   - Numerical results

Experimental setup: Quadratic sensing

Quadratic Sensing [CCG15]: recover a rank $r_\natural = 3$ matrix $U_\natural \in \mathbf{R}^{n\times r_\natural}$ with $\|U_\natural\|_{\mathsf{F}}^2 = 1$ from quadratic measurement $y \in \mathbf{R}^m$

Experimental setup: Quadratic sensing

Quadratic Sensing [CCG15]: recover a rank $r_\natural = 3$ matrix $U_\natural \in \mathbf{R}^{n\times r_\natural}$ with $\|U_\natural\|_F^2 = 1$ from quadratic measurement $y \in \mathbf{R}^m$

1. random standard gaussian measurements a_i
2. $y_0(i) = \left\|U_\natural^\top a_i\right\|_F^2, i = 1, \dots, m,\ m = 15nr_\natural$

Experimental setup: Quadratic sensing

Quadratic Sensing [CCG15]: recover a rank $r_\natural = 3$ matrix $U_\natural \in \mathbf{R}^{n \times r_\natural}$ with $\|U_\natural\|_F^2 = 1$ from quadratic measurement $y \in \mathbf{R}^m$

1. random standard gaussian measurements a_i
2. $y_0(i) = \left\| U_\natural^\top a_i \right\|_F^2, i = 1, \dots, m, \; m = 15nr_\natural$
3. $y = y_0 + c \|y_0\|_2 v$, c is the inverse signal-to-noise ratio, v is a random unit vector

Experimental setup: Quadratic sensing

Quadratic Sensing [CCG15]: recover a rank $r_\natural = 3$ matrix $U_\natural \in \mathbf{R}^{n\times r_\natural}$ with $\|U_\natural\|_F^2 = 1$ from quadratic measurement $y \in \mathbf{R}^m$

1. random standard gaussian measurements a_i
2. $y_0(i) = \left\|U_\natural^\top a_i\right\|_F^2, i = 1, \ldots, m,\ m = 15nr_\natural$
3. $y = y_0 + c\,\|y_0\|_2\, v$, c is the inverse signal-to-noise ratio, v is a random unit vector

Optimization problem:

$$\begin{array}{ll} \text{minimize} & f(X) := \frac{1}{2}\sum_{i=1}^{m}\left(a_i^\top X a_i - y_i\right)^2 \\ \text{subject to} & \mathbf{tr}(X) = \tau, \quad X \succeq 0. \end{array} \qquad \text{(Quadratic Sensing)}$$

Set $\tau = \frac{1}{2}$ and $c = 0.5$ in numerics.

Low rank solution and strict complementarity

Dimension n	Avg. gap	Avg. recovery error
100	288.06	0.0013
200	505.16	0.00064
400	961.09	0.00031
600	1358.62	0.00021

Table: Verification of low rankness and strict complementarity. Rank $r_\star = 3$ in all experiments. The recovery error is measured by $\frac{\left\|\frac{X_\star}{\tau} - U_\natural U_\natural^\top\right\|_{\mathrm{F}}}{\left\|U_\natural U_\natural^\top\right\|_{\mathrm{F}}}$. The gap is measured by $\lambda_{n-3}(\nabla f(X_\star)) - \lambda_n(\nabla f(X_\star))$. All the results are averaged over 20 iid trials.

Numerical results $k > r_\star$

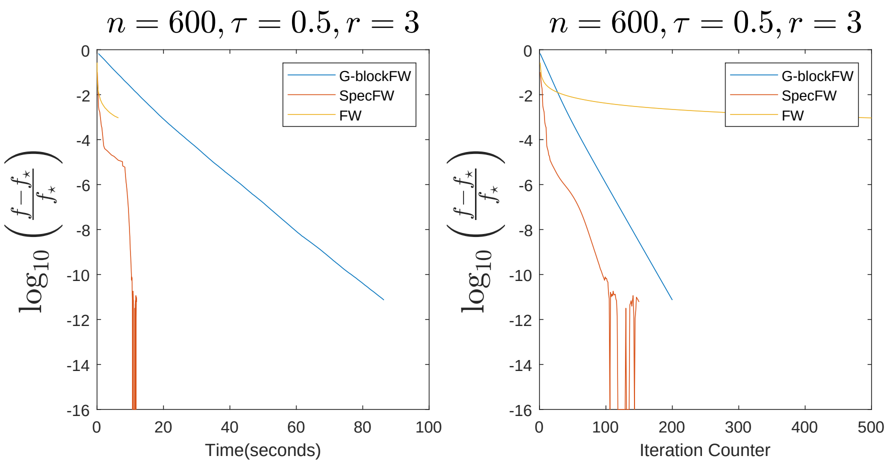

Figure: $k > r_\star$. comparison of algorithms FW, G-blockFW [AZHHL17], and SpecFW. Left: accuracy vs time. Right: accuracy vs iteration.

Numerical results $k < r_\star$

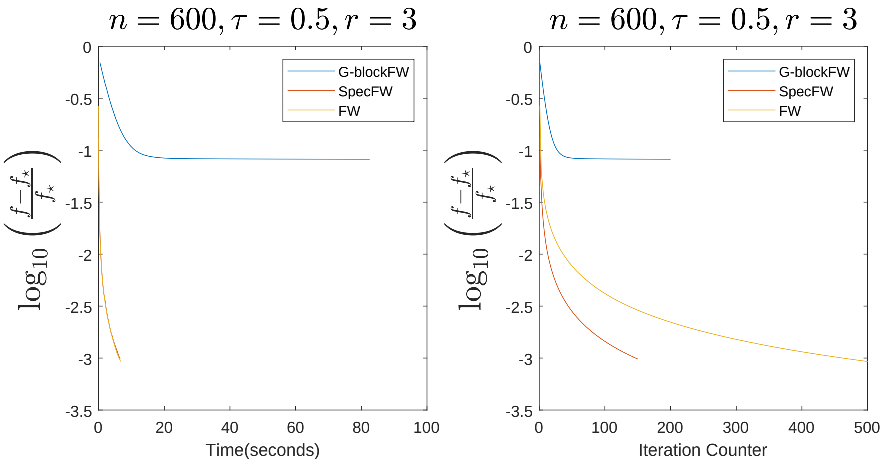

Figure: $k < r_\star$. comparison of algorithms FW, G-blockFW [AZHHL17], and SpecFW. Left: accuracy vs time. Right: accuracy vs iteration.

References I

Ali Ahmed, Benjamin Recht, and Justin Romberg.
Blind deconvolution using convex programming.
IEEE Transactions on Information Theory, 60(3):1711–1732, 2013.

Zeyuan Allen-Zhu, Elad Hazan, Wei Hu, and Yuanzhi Li.
Linear convergence of a frank-wolfe type algorithm over trace-norm balls.
In *Advances in Neural Information Processing Systems*, pages 6191–6200, 2017.

Yuxin Chen, Yuejie Chi, and Andrea J Goldsmith.
Exact and stable covariance estimation from quadratic sampling via convex programming.
IEEE Transactions on Information Theory, 61(7):4034–4059, 2015.

References II

Emmanuel J Candes, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski.
Phase retrieval via matrix completion.
SIAM review, 57(2):225–251, 2015.

Emmanuel J Candès and Benjamin Recht.
Exact matrix completion via convex optimization.
Foundations of Computational mathematics, 9(6):717, 2009.

Mark A Davenport, Yaniv Plan, Ewout Van Den Berg, and Mary Wootters.
1-bit matrix completion.
Information and Inference: A Journal of the IMA, 3(3):189–223, 2014.

References III

Robert M Freund, Paul Grigas, and Rahul Mazumder.
An extended frank-wolfe method with “in-face” directions, and its application to low-rank matrix completion.
SIAM Journal on optimization, 27(1):319–346, 2017.

Dan Garber.
Faster projection-free convex optimization over the spectrahedron.
In *Advances in Neural Information Processing Systems*, pages 874–882, 2016.

Dan Garber.
Linear convergence of frank-wolfe for rank-one matrix recovery without strong convexity.
arXiv preprint arXiv:1912.01467, 2019.

References IV

Martin Jaggi and Marek Sulovskỳ.
A simple algorithm for nuclear norm regularized problems.
2010.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo.
Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization.
SIAM review, 52(3):471–501, 2010.

Alp Yurtsever, Madeleine Udell, Joel Tropp, and Volkan Cevher.
Sketchy decisions: Convex low-rank matrix optimization with optimal storage.
In *Artificial Intelligence and Statistics*, pages 1188–1196, 2017.