# Multiplicative Gradient Method: When and Why It Works

**Renbo Zhao**

MIT Operations Research Center

24th Midwest Optimization Meeting
University of Waterloo
October, 2022

# Positron Emission Tomography

$$\max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \tag{PET}$$

# Positron Emission Tomography

$$\max_{x \in \Delta_n} \left\{ F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \qquad \text{(PET)}$$

- Known as Positron Emission Tomography (PET) in medical imaging, but has many other applications, e.g., inference of multi-dimensional Hawkes processes [ZZS13] and log-optimal investment [Cov84].

# Positron Emission Tomography

$$\max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \tag{PET}$$

- Known as Positron Emission Tomography (PET) in medical imaging, but has many other applications, e.g., inference of multi-dimensional Hawkes processes [ZZS13] and log-optimal investment [Cov84].
- For all $j \in [m]$, let $p_j > 0$ and $\sum_{j=1}^m p_j = 1$.

# Positron Emission Tomography

$$\max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \tag{PET}$$

- Known as Positron Emission Tomography (PET) in medical imaging, but has many other applications, e.g., inference of multi-dimensional Hawkes processes [ZZS13] and log-optimal investment [Cov84].
- For all $j \in [m]$, let $p_j > 0$ and $\sum_{j=1}^m p_j = 1$.
- For all $j \in [m]$, $a_j \neq 0$, $a_j \in \mathbb{R}^n_+$ $(a_j \geq 0)$ and can be "sparse".

# Positron Emission Tomography

$$\max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \qquad \text{(PET)}$$

- Known as Positron Emission Tomography (PET) in medical imaging, but has many other applications, e.g., inference of multi-dimensional Hawkes processes [ZZS13] and log-optimal investment [Cov84].
- For all $j \in [m]$, let $p_j > 0$ and $\sum_{j=1}^m p_j = 1$.
- For all $j \in [m]$, $a_j \neq 0$, $a_j \in \mathbb{R}_+^n$ $(a_j \geq 0)$ and can be "sparse".
- $\Delta_n := \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1\}$ is the unit simplex in $\mathbb{R}^n$.

# Positron Emission Tomography

$$\max_{x\in\Delta_n} \left\{F(x) := \textstyle\sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \qquad \text{(PET)}$$

- Known as Positron Emission Tomography (PET) in medical imaging, but has many other applications, e.g., inference of multi-dimensional Hawkes processes [ZZS13] and log-optimal investment [Cov84].
- For all $j \in [m]$, let $p_j > 0$ and $\sum_{j=1}^m p_j = 1$.
- For all $j \in [m]$, $a_j \neq 0$, $a_j \in \mathbb{R}^n_+$ ($a_j \geq 0$) and can be "sparse".
- $\Delta_n := \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ is the unit simplex in $\mathbb{R}^n$.
- Multiplicative gradient method: $x^0 \in \mathsf{ri}\,\Delta_n$

$$x^{t+1} = x^t \circ \nabla F(x^t) \quad \equiv \quad x_i^{t+1} := x_i^t \nabla_i F(x^t), \quad \forall\, i \in [n].$$

# Positron Emission Tomography

$$\max_{x\in\Delta_n} \left\{F(x) := \textstyle\sum_{j=1}^{m} p_j \ln(a_j^\top x)\right\} \qquad \text{(PET)}$$

- Known as Positron Emission Tomography (PET) in medical imaging, but has many other applications, e.g., inference of multi-dimensional Hawkes processes [ZZS13] and log-optimal investment [Cov84].
- For all $j \in [m]$, let $p_j > 0$ and $\sum_{j=1}^{m} p_j = 1$.
- For all $j \in [m]$, $a_j \neq 0$, $a_j \in \mathbb{R}_+^n$ ($a_j \geq 0$) and can be "sparse".
- $\Delta_n := \{x \in \mathbb{R}_+^n : \sum_{i=1}^{n} x_i = 1\}$ is the unit simplex in $\mathbb{R}^n$.
- Multiplicative gradient method: $x^0 \in \mathsf{ri}\,\Delta_n$

$$x^{t+1} = x^t \circ \nabla F(x^t) \quad \equiv \quad x_i^{t+1} := x_i^t \nabla_i F(x^t), \quad \forall\, i \in [n]. \qquad \text{(MG)}$$

- (MG) does not fall under any "well-known" optimization frameworks, e.g., Newton-type method, mirror descent, etc.

# The Mystery of MG

$$\max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \tag{PET}$$

$$\boxed{x^0 \in \mathsf{ri}\,\Delta_n, \qquad x^{t+1} = x^t \circ \nabla F(x^t)} \tag{MG}$$

# The Mystery of MG

$$\max_{x \in \Delta_n} \; \left\{ F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \tag{PET}$$

$$\boxed{x^0 \in \mathsf{ri}\,\Delta_n, \qquad x^{t+1} = x^t \circ \nabla F(x^t)} \tag{MG}$$

▷ Originally proposed by information theorists in the 1970s [Ari72] based on the EM procedure.

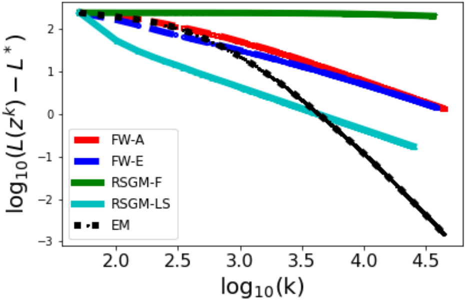
# The Mystery of MG

$$\max_{x \in \Delta_n} \left\{ F(x) := \textstyle\sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \qquad \text{(PET)}$$

$$\boxed{x^0 \in \mathsf{ri}\,\Delta_n, \qquad x^{t+1} = x^t \circ \nabla F(x^t)} \qquad \text{(MG)}$$

- Originally proposed by information theorists in the 1970s [Ari72] based on the EM procedure.

- Impressive numerical performance: $x^0 = (1/n)e$




FW-A & FW-E [Dvu20; ZF22]: Frank-Wolfe (FW) method for logarithmically-homogeneous self-concordant barriers (with adaptive stepsize and exact line search)

RSGM-F & RSGM-LS [BBT17; LFN18]: Relatively smooth gradient method (with fixed stepsize and backtracking line search)

# The Mystery of MG

$$F^* = \max_{x \in \Delta_n} \left\{ F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \qquad \text{(PET)}$$

- However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.

# The Mystery of MG

$$F^* = \max_{x \in \Delta_n} \left\{ F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \qquad \text{(PET)}$$

- However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- Zhao [Zha21] showed that (MG) has the following convergence rate:

$$F^* - F(x^t) \le \ln(n)/t, \qquad \forall t \ge 1.$$

# The Mystery of MG

$$F^* = \max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \qquad \text{(PET)}$$

- However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.

- Zhao [Zha21] showed that (MG) has the following convergence rate:

$$F^* - F(x^t) \le \ln(n)/t, \qquad \forall t \ge 1.$$

The proof is relatively short, and is based on basic convex analysis.

# The Mystery of MG

$$F^* = \max_{x \in \Delta_n} \left\{ F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \qquad \text{(PET)}$$

- However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- Zhao [Zha21] showed that (MG) has the following convergence rate:

  $$F^* - F(x^t) \le \ln(n)/t, \qquad \forall t \ge 1.$$

  The proof is relatively short, and is based on basic convex analysis.
- But immediately some questions arose:

# The Mystery of MG

$$F^* = \max_{x \in \Delta_n} \left\{ F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x) \right\} \qquad \text{(PET)}$$

▷ However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.

▷ Zhao [Zha21] showed that (MG) has the following convergence rate:

$$F^* - F(x^t) \le \ln(n)/t, \qquad \forall t \ge 1.$$

The proof is relatively short, and is based on basic convex analysis.

▷ But immediately some questions arose:
- Why does (MG) work for PET?

# The Mystery of MG

$$F^* = \max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \qquad \text{(PET)}$$

- ▷ However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- ▷ Zhao [Zha21] showed that (MG) has the following convergence rate:

  $$F^* - F(x^t) \le \ln(n)/t, \qquad \forall t \ge 1.$$

  The proof is relatively short, and is based on basic convex analysis.
- ▷ But immediately some questions arose:
  - Why does (MG) work for PET?
  - What are the essential structures of the problem the drive the success of (MG)? Is there a general problem class that (MG) works well?

# The Mystery of MG

$$F^* = \max_{x\in\Delta_n} \left\{F(x) := \sum_{j=1}^m p_j \ln(a_j^\top x)\right\} \tag{PET}$$

- ▷ However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- ▷ Zhao [Zha21] showed that (MG) has the following convergence rate:

  $$F^* - F(x^t) \le \ln(n)/t, \qquad \forall t \ge 1.$$

  The proof is relatively short, and is based on basic convex analysis.
- ▷ But immediately some questions arose:
  - Why does (MG) work for PET?
  - What are the essential structures of the problem the drive the success of (MG)? Is there a general problem class that (MG) works well?
  - And what is the interaction between the complexity of (MG) and the problem structure?

# Our Main Contributions

- We identify a broad problem class and develop a generalization of the MG method, with computational guarantees.

# Our Main Contributions

- ▷ We identify a broad problem class and develop a generalization of the MG method, with computational guarantees.
  - The problem class is best viewed via the notion of symmetric cones.

# Our Main Contributions

- ▷ We identify a broad problem class and develop a generalization of the MG method, with computational guarantees.
  - The problem class is best viewed via the notion of symmetric cones.
  - The development and analysis of the generalized MG method are based on the framework of Euclidean Jordan algebra.

# Our Main Contributions

- We identify a broad problem class and develop a generalization of the MG method, with computational guarantees.
  - The problem class is best viewed via the notion of symmetric cones.
  - The development and analysis of the generalized MG method are based on the framework of Euclidean Jordan algebra.
- To facilitate understanding, we first show our results when they are specialized to the following applications:
  - D-optimal design
  - Quantum state tomography
  - Semidefinite relaxation of Boolean QP

# Our Main Contributions

- We identify a broad problem class and develop a generalization of the MG method, with computational guarantees.
  - The problem class is best viewed via the notion of symmetric cones.
  - The development and analysis of the generalized MG method are based on the framework of Euclidean Jordan algebra.
- To facilitate understanding, we first show our results when they are specialized to the following applications:
  - D-optimal design
  - Quantum state tomography
  - Semidefinite relaxation of Boolean QP
- In all of these applications, the objective functions involve "$\ln(\cdot)$", and hence are neither Lipschitz nor smooth (i.e., have Lipschitz gradients) on the feasible sets.

# Our Main Contributions

- ▷ We identify a broad problem class and develop a generalization of the MG method, with computational guarantees.
  - The problem class is best viewed via the notion of symmetric cones.
  - The development and analysis of the generalized MG method are based on the framework of Euclidean Jordan algebra.
- ▷ To facilitate understanding, we first show our results when they are specialized to the following applications:
  - D-optimal design
  - Quantum state tomography
  - Semidefinite relaxation of Boolean QP
- ▷ In all of these applications, the objective functions involve "$\ln(\cdot)$", and hence are neither Lipschitz nor smooth (i.e., have Lipschitz gradients) on the feasible sets.
- ▷ Certain first-order methods for these applications have been developed recently [Nes11; BBT17; LFN18; Dvu20; ZF22] — our generalized MG method contributes to this line of research from a different viewpoint.

# D-Optimal Design (D-OPT)

$$\max_x F(x) := m^{-1} \ln\det\left(\textstyle\sum_{i=1}^n x_i a_i a_i^\top\right) \quad \text{s.t.} \quad x \in \Delta_n \qquad \text{(D-OPT)}$$

# D-Optimal Design (D-OPT)

$$\max_x F(x) := m^{-1} \ln\det\left(\textstyle\sum_{i=1}^n x_i a_i a_i^\top\right) \quad \text{s.t.} \quad x \in \Delta_n \qquad \text{(D-OPT)}$$

▷ $a_1, \ldots, a_n \in \mathbb{R}^m$ whose linear span is $\mathbb{R}^m$.

# D-Optimal Design (D-OPT)

$$\max_x F(x) := m^{-1} \ln\det\left(\textstyle\sum_{i=1}^n x_i a_i a_i^\top\right) \quad \text{s.t.} \quad x \in \Delta_n \tag{D-OPT}$$

- $a_1, \ldots, a_n \in \mathbb{R}^m$ whose linear span is $\mathbb{R}^m$.
- In statistics, D-OPT corresponds to maximizing the the determinant of the Fisher information matrix [Fed72].

# D-Optimal Design (D-OPT)

$$\max_x F(x) := m^{-1} \ln\det\left(\textstyle\sum_{i=1}^n x_i a_i a_i^\top\right) \quad \text{s.t.} \quad x \in \Delta_n \qquad \text{(D-OPT)}$$

- $a_1, \ldots, a_n \in \mathbb{R}^m$ whose linear span is $\mathbb{R}^m$.
- In statistics, D-OPT corresponds to maximizing the the determinant of the Fisher information matrix [Fed72].
- In computational geometry, D-OPT arises as a Lagrangian dual problem of the minimum volume enclosing ellipsoid problem [Tod16].

# D-Optimal Design (D-OPT)

$$\max_x F(x) := m^{-1} \ln\det\left(\sum_{i=1}^n x_i a_i a_i^\top\right) \quad \text{s.t.} \quad x \in \Delta_n \tag{D-OPT}$$

- $a_1, \ldots, a_n \in \mathbb{R}^m$ whose linear span is $\mathbb{R}^m$.
- In statistics, D-OPT corresponds to maximizing the the determinant of the Fisher information matrix [Fed72].
- In computational geometry, D-OPT arises as a Lagrangian dual problem of the minimum volume enclosing ellipsoid problem [Tod16].
- MG method: $\boxed{x^0 \in \mathsf{ri}\,\Delta_n, \quad x^{t+1} = x^t \circ \nabla F(x^t)}$

# Quantum State Tomography (QST)

$$\begin{aligned} \max_X \; & F(X) := m^{-1} \textstyle\sum_{j=1}^q n_j \ln(\langle X, a_j a_j^H \rangle) \\ \text{s. t.} \quad & X \in \mathbb{H}_+^n, \;\; \mathrm{tr}(X) = \langle I_n, X \rangle = 1 \end{aligned} \tag{QST}$$

# Quantum State Tomography (QST)

$$\begin{aligned} &\max_X \; F(X) := m^{-1} \textstyle\sum_{j=1}^q n_j \ln(\langle X, a_j a_j^H \rangle) \\ &\text{s. t.} \quad X \in \mathbb{H}_+^n, \;\; \mathrm{tr}(X) = \langle I_n, X \rangle = 1 \end{aligned} \tag{QST}$$

- In quantum physics, this problem aims to reconstruct the state of a quantum system using the measured output of particles [Hra04].

# Quantum State Tomography (QST)

$$\begin{aligned} \max_X \; & F(X) := m^{-1} \textstyle\sum_{j=1}^q n_j \ln(\langle X, a_j a_j^H \rangle) \\ \text{s.t.} \quad & X \in \mathbb{H}_+^n, \;\; \mathrm{tr}(X) = \langle I_n, X \rangle = 1 \end{aligned} \tag{QST}$$

- In quantum physics, this problem aims to reconstruct the state of a quantum system using the measured output of particles [Hra04].
- $a_1, \ldots, a_q \in \mathbb{C}^n$, $\sum_{j=1}^q a_j a_j^H = I_n$ and $\sum_{j=1}^q n_j = m$.

# Quantum State Tomography (QST)

$$\begin{aligned} \max_X \; & F(X) := m^{-1} \textstyle\sum_{j=1}^q n_j \ln(\langle X, a_j a_j^H \rangle) \\ \text{s.t.} \;\; & X \in \mathbb{H}_+^n, \;\; \mathrm{tr}(X) = \langle I_n, X \rangle = 1 \end{aligned} \tag{QST}$$

- In quantum physics, this problem aims to reconstruct the state of a quantum system using the measured output of particles [Hra04].
- $a_1, \ldots, a_q \in \mathbb{C}^n$, $\sum_{j=1}^q a_j a_j^H = I_n$ and $\sum_{j=1}^q n_j = m$.
- $\mathbb{H}_+^n$ denotes the cone of $n \times n$ complex Hermitian PSD matrices.

# Quantum State Tomography (QST)

$$\begin{aligned} &\max_X \ F(X) := m^{-1} \textstyle\sum_{j=1}^q n_j \ln(\langle X, a_j a_j^H \rangle) \\ &\text{s.t.} \quad X \in \mathbb{H}_+^n, \ \ \mathrm{tr}(X) = \langle I_n, X \rangle = 1 \end{aligned} \tag{QST}$$

- In quantum physics, this problem aims to reconstruct the state of a quantum system using the measured output of particles [Hra04].
- $a_1, \ldots, a_q \in \mathbb{C}^n$, $\sum_{j=1}^q a_j a_j^H = I_n$ and $\sum_{j=1}^q n_j = m$.
- $\mathbb{H}_+^n$ denotes the cone of $n \times n$ complex Hermitian PSD matrices.
- (Generalized) MG method: $X^0 \succ 0$, $\mathrm{tr}(X^0) = 1$,

$$\boxed{\begin{aligned} \hat{X}^{t+1} &= \exp\{\ln(X^t) + \ln(\nabla F(X^t))\} \\ X^{t+1} &= \hat{X}^{t+1} / \mathrm{tr}(\hat{X}^{t+1}) \end{aligned}}$$

(For any $X = \sum_{i=1}^n \lambda_i u_i u_i^H \succ 0$, $\ln(X) := \ln(\lambda_i) u_i u_i^H$.)

# Semidefinite Relaxation of Boolean QP (RBQP)

- ▷ The Boolean QP (BQP): $q^* := \max_{x\in\{\pm 1\}^n} x^\top Ax$ for some $A \succ 0$.

# Semidefinite Relaxation of Boolean QP (RBQP)

- The Boolean QP (BQP): $q^* := \max_{x\in\{\pm 1\}^n} x^\top A x$ for some $A \succ 0$.
- Nesterov [Nes98] showed that the semidefinite relaxation

$$s^* := \min_y \ \langle e, y\rangle \quad \text{s.t.} \quad \mathrm{Diag}(y) \succeq A \tag{SDP}$$

provides a $(2/\pi)$-approximation of the BQP.

# Semidefinite Relaxation of Boolean QP (RBQP)

- ▷ The Boolean QP (BQP): $q^* := \max_{x\in\{\pm 1\}^n} x^\top A x$ for some $A \succ 0$.
- ▷ Nesterov [Nes98] showed that the semidefinite relaxation

$$s^* := \min_y \ \langle e, y\rangle \quad \text{s.t.} \quad \mathrm{Diag}(y) \succeq A \tag{SDP}$$

provides a $(2/\pi)$-approximation of the BQP.

- ▷ Nesterov [Nes11] later showed that (SDP) above can be equivalently written in the dual form:

$$\begin{aligned} \max_X \quad & F(X) := 2\ln\left(\textstyle\sum_{i=1}^n \langle X, r_i r_i^\top\rangle^{1/2}\right) \\ \text{s.t.} \quad & X \in \mathbb{S}_+^n,\ \langle I_n, X\rangle = 1 \end{aligned} \tag{RBQP}$$

where $A = R^\top R$ and $R := [r_1 \ \cdots \ r_n]$, and $\mathbb{S}_+^n$ denotes the cone of $n \times n$ real symmetric PSD matrices.

# Semidefinite Relaxation of Boolean QP (RBQP)

- The Boolean QP (BQP): $q^* := \max_{x\in\{\pm1\}^n} x^\top A x$ for some $A \succ 0$.
- Nesterov [Nes98] showed that the semidefinite relaxation

$$s^* := \min_y \ \langle e, y\rangle \quad \text{s.t.} \quad \mathrm{Diag}(y) \succeq A \tag{SDP}$$

  provides a $(2/\pi)$-approximation of the BQP.
- Nesterov [Nes11] later showed that (SDP) above can be equivalently written in the dual form:

$$\begin{aligned} \max_X \quad & F(X) := 2\ln\left(\textstyle\sum_{i=1}^n \langle X, r_i r_i^\top\rangle^{1/2}\right) \\ \text{s.t.} \quad & X \in \mathbb{S}_+^n,\ \langle I_n, X\rangle = 1 \end{aligned} \tag{RBQP}$$

  where $A = R^\top R$ and $R := [r_1 \ \cdots \ r_n]$, and $\mathbb{S}_+^n$ denotes the cone of $n \times n$ real symmetric PSD matrices.
- MG method: $X^0 \succ 0$, $\mathrm{tr}(X^0) = 1$,

$$\boxed{\begin{aligned} \hat{X}^{t+1} &= \exp\{\ln(X^t) + \ln(\nabla F(X^t))\} \\ X^{t+1} &= \hat{X}^{t+1} / \mathrm{tr}(\hat{X}^{t+1}) \end{aligned}}$$

# Comparison of Computational Guarantees

RSGM [BBT17; LFN18]: Relatively smooth gradient method
FW [Dvu20; ZF22]: FW method for logarithmically-homogeneous self-concordant barriers
MG: (Generalized) Multiplicative gradient method (this work)
BSG [Nes11]: Barrier subgradient method

Table 1: Comparison of operations complexities (with $x^0 = (1/n)e$ or $X^0 = (1/n)I_n$)

| | RSGM | FW | MG | BSG | Regime |
|---|---|---|---|---|---|
| PET | $O\left(\frac{mn^2}{\varepsilon}\ln\left(\frac{\ln(n)}{\varepsilon}\right)\right)$ | $O\left(\frac{m^2 n}{\varepsilon}\right)$ | $O\left(\frac{mn\ln(n)}{\varepsilon}\right)$ | $O\left(\frac{mn^2}{\varepsilon^2}\ln^2\left(\frac{n}{\varepsilon}\right)\right)$ | $n = O(\exp(m))$ |
| D-OPT | $O\left(\frac{mn^2}{\varepsilon}\ln\left(\frac{\ln(n/m)}{\varepsilon}\right)\right)$ | $O\left(\frac{m^2 n}{\varepsilon}\right)$ | $O\left(\frac{m^2 n\ln(n)}{\varepsilon}\right)^\dagger$ | $O\left(\frac{m^2 n^2}{\varepsilon^2}\ln^2\left(\frac{n}{\varepsilon}\right)\right)$ | |
| QST | x? | $O\left(\frac{m^2 n^2}{\varepsilon}\right)$ | $O\left(\frac{mn^2\ln(n)}{\varepsilon}\right)^\ddagger$ | $O\left(\frac{mn^3}{\varepsilon^2}\ln^2\left(\frac{n}{\varepsilon}\right)\right)$ | $n = O(\exp(m))$ |
| RBQP | x? | x? | $O\left(\frac{n^3\ln(n)}{\varepsilon}\right)$ | $O\left(\frac{n^4}{\varepsilon^2}\ln^2\left(\frac{n}{\varepsilon}\right)\right)$ | |

† [Coh19] ‡ [LCL21]

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

▷ $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.

# A General Problem Class

$$\begin{aligned} \max\ & F(x) := f(\mathsf{A}x) \\ \text{s.t.}\ & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.
- $e \in \mathsf{int}\,\mathcal{K}_1$ is the "center" of $\mathcal{K}_1$, e.g., $e = \mathbf{1}_n := (1, \ldots, 1)$ if $\mathcal{K}_1 = \mathbb{R}^n_+$ and $e = I_n$ if $\mathcal{K}_1 = \mathbb{S}^n_+$.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- ▷ $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.
- ▷ $e \in \mathsf{int}\,\mathcal{K}_1$ is the "center" of $\mathcal{K}_1$, e.g., $e = \mathbf{1}_n := (1, \ldots, 1)$ if $\mathcal{K}_1 = \mathbb{R}^n_+$ and $e = I_n$ if $\mathcal{K}_1 = \mathbb{S}^n_+$.
- ▷ $\mathsf{A} : \mathcal{K}_1 \to \mathcal{K}_2$ is a linear operator, where $\mathcal{K}_2$ is any regular cone.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.
- $e \in \mathsf{int}\,\mathcal{K}_1$ is the "center" of $\mathcal{K}_1$, e.g., $e = \mathbf{1}_n := (1, \ldots, 1)$ if $\mathcal{K}_1 = \mathbb{R}^n_+$ and $e = I_n$ if $\mathcal{K}_1 = \mathbb{S}^n_+$.
- $\mathsf{A} : \mathcal{K}_1 \to \mathcal{K}_2$ is a linear operator, where $\mathcal{K}_2$ is any regular cone.
  - We require both $\mathsf{A} : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_2$ and $\mathsf{A}^* : \mathsf{int}\,\mathcal{K}_2^* \to \mathsf{int}\,\mathcal{K}_1$.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- ▷ $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.
- ▷ $e \in \operatorname{int} \mathcal{K}_1$ is the "center" of $\mathcal{K}_1$, e.g., $e = \mathbf{1}_n := (1, \ldots, 1)$ if $\mathcal{K}_1 = \mathbb{R}^n_+$ and $e = I_n$ if $\mathcal{K}_1 = \mathbb{S}^n_+$.
- ▷ $\mathsf{A} : \mathcal{K}_1 \to \mathcal{K}_2$ is a linear operator, where $\mathcal{K}_2$ is any regular cone.
  - We require both $\mathsf{A} : \operatorname{int} \mathcal{K}_1 \to \operatorname{int} \mathcal{K}_2$ and $\mathsf{A}^* : \operatorname{int} \mathcal{K}_2^* \to \operatorname{int} \mathcal{K}_1$.
- ▷ $f : \mathcal{K}_2 \to \mathbb{R} \cup \{-\infty\}$ is concave, and is three-times differentiable and 1-logarithmically-homogeneous (1-LH) on $\operatorname{int} \mathcal{K}_2$, namely

$$f(ty) = f(y) + \ln t, \quad \forall\, y \in \operatorname{int} \mathcal{K}_2,\ \forall\, t > 0.$$

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s. t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- ▷ $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.
- ▷ $e \in \mathsf{int}\,\mathcal{K}_1$ is the "center" of $\mathcal{K}_1$, e.g., $e = \mathbf{1}_n := (1, \ldots, 1)$ if $\mathcal{K}_1 = \mathbb{R}^n_+$ and $e = I_n$ if $\mathcal{K}_1 = \mathbb{S}^n_+$.
- ▷ $\mathsf{A} : \mathcal{K}_1 \to \mathcal{K}_2$ is a linear operator, where $\mathcal{K}_2$ is any regular cone.
  - We require both $\mathsf{A} : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_2$ and $\mathsf{A}^* : \mathsf{int}\,\mathcal{K}_2^* \to \mathsf{int}\,\mathcal{K}_1$.
- ▷ $f : \mathcal{K}_2 \to \mathbb{R} \cup \{-\infty\}$ is concave, and is three-times differentiable and 1-logarithmically-homogeneous (1-LH) on $\mathsf{int}\,\mathcal{K}_2$, namely

$$f(ty) = f(y) + \ln t, \quad \forall\, y \in \mathsf{int}\,\mathcal{K}_2,\ \forall\, t > 0.$$

  Note that $-f$ is not necessarily a self-concordant function (or barrier).
  - Consider $f(x) = 2\ln(\sum_{i=1}^n \sqrt{x_i})$ for $x \in \mathbb{R}^n_+ \setminus \{0\}$ in RBQP.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s. t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x\rangle = 1\} \end{aligned} \tag{P}$$

- $\mathcal{K}_1$ is a symmetric cone (self-dual and homogeneous) with rank $n$.
- $e \in \mathsf{int}\,\mathcal{K}_1$ is the "center" of $\mathcal{K}_1$, e.g., $e = \mathbf{1}_n := (1, \ldots, 1)$ if $\mathcal{K}_1 = \mathbb{R}^n_+$ and $e = I_n$ if $\mathcal{K}_1 = \mathbb{S}^n_+$.
- $\mathsf{A} : \mathcal{K}_1 \to \mathcal{K}_2$ is a linear operator, where $\mathcal{K}_2$ is any regular cone.
  - We require both $\mathsf{A} : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_2$ and $\mathsf{A}^* : \mathsf{int}\,\mathcal{K}_2^* \to \mathsf{int}\,\mathcal{K}_1$.
- $f : \mathcal{K}_2 \to \mathbb{R} \cup \{-\infty\}$ is concave, and is three-times differentiable and 1-logarithmically-homogeneous (1-LH) on $\mathsf{int}\,\mathcal{K}_2$, namely

  $$f(ty) = f(y) + \ln t, \quad \forall\, y \in \mathsf{int}\,\mathcal{K}_2,\ \forall\, t > 0.$$

  Note that $-f$ is not necessarily a self-concordant function (or barrier).
  - Consider $f(x) = 2\ln(\sum_{i=1}^n \sqrt{x_i})$ for $x \in \mathbb{R}^n_+ \setminus \{0\}$ in RBQP.
- We require $F$ to lie in the class of *gradient log-convex* functions.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a (finite-dimensional) inner-product space.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a (finite-dimensional) inner-product space.
- $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is
  - self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
  - homogeneous: for all $x, y \in \operatorname{int} \mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a (finite-dimensional) inner-product space.
- $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is
  - self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
  - homogeneous: for all $x, y \in \operatorname{int} \mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$
- A symmetric cone is either one of the following five *primitive* cones:

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a (finite-dimensional) inner-product space.
- $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is
  - self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
  - homogeneous: for all $x, y \in \mathsf{int}\,\mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$
- A symmetric cone is either one of the following five *primitive* cones:
  - the second-order cone $\mathcal{Q}^{n+1}$,

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

▷ $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a (finite-dimensional) inner-product space.

▷ $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is

- self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
- homogeneous: for all $x, y \in \operatorname{int} \mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$

▷ A symmetric cone is either one of the following five *primitive* cones:

- the second-order cone $\mathcal{Q}^{n+1}$,
- the cone of $n \times n$ Hermitian PSD matrices over reals, complexes and quaternions,

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x\rangle = 1\} \end{aligned} \tag{P}$$

▷ $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ is a (finite-dimensional) inner-product space.

▷ $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is
- self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
- homogeneous: for all $x, y \in \mathsf{int}\,\mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$

▷ A symmetric cone is either one of the following five *primitive* cones:
- the second-order cone $\mathcal{Q}^{n+1}$,
- the cone of $n \times n$ Hermitian PSD matrices over reals, complexes and quaternions,
- the cone of $3 \times 3$ Hermitian PSD matrices over octonions (a.k.a. the 27-dimensional exceptional cone),

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x\rangle = 1\} \end{aligned} \tag{P}$$

- ▷ $(\mathbb{V}, \langle\cdot,\cdot\rangle)$ is a (finite-dimensional) inner-product space.
- ▷ $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is
  - self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
  - homogeneous: for all $x, y \in \mathsf{int}\,\mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$
- ▷ A symmetric cone is either one of the following five *primitive* cones:
  - the second-order cone $\mathcal{Q}^{n+1}$,
  - the cone of $n \times n$ Hermitian PSD matrices over reals, complexes and quaternions,
  - the cone of $3 \times 3$ Hermitian PSD matrices over octonions (a.k.a. the 27-dimensional exceptional cone),

  or the Cartesian product of these primitive cones, e.g.,$\mathbb{R}^m_+$.

# A General Problem Class

$$\begin{aligned} \max \quad & F(x) := f(\mathsf{A}x) \\ \text{s.t.} \quad & x \in \mathcal{C} := \{x \in \mathcal{K}_1 : \langle e, x \rangle = 1\} \end{aligned} \tag{P}$$

- $(\mathbb{V}, \langle \cdot, \cdot \rangle)$ is a (finite-dimensional) inner-product space.
- $\mathcal{K}_1 \subseteq \mathbb{V}$ is a symmetric cone if it is
  - self-dual: $\mathcal{K}_1 = \mathcal{K}_1^*$ (the dual cone of $\mathcal{K}_1$)
  - homogeneous: for all $x, y \in \mathsf{int}\, \mathcal{K}_1$, there exists a linear automorphism $\mathsf{T}$ on $\mathcal{K}_1$ such that $\mathsf{T}x = y$
- A symmetric cone is either one of the following five *primitive* cones:
  - the second-order cone $\mathcal{Q}^{n+1}$,
  - the cone of $n \times n$ Hermitian PSD matrices over reals, complexes and quaternions,
  - the cone of $3 \times 3$ Hermitian PSD matrices over octonions (a.k.a. the 27-dimensional exceptional cone),

  or the Cartesian product of these primitive cones, e.g.,$\mathbb{R}^m_+$.
- $\mathcal{C}$ is sometimes referred to as the “generalized unit simplex”, including unit simplex, unit $\ell_2$-ball and spectrahedron.

# The Class of Gradient Log-Convex Functions

We call $F$ in (P) *gradient log-convex* if $\nabla F : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_1$ satisfies

$$\boxed{\begin{aligned} \ln(\nabla F(\lambda x + (1-\lambda)y)) \preceq_{\mathcal{K}_1} \lambda \ln(\nabla F(x)) + (1-\lambda)\ln(\nabla F(y)) \\ \forall x, y \in \mathsf{int}\,\mathcal{K}_1, \;\; \forall\, \lambda \in [0,1] \end{aligned}} \tag{GLC}$$

# The Class of Gradient Log-Convex Functions

We call $F$ in (P) *gradient log-convex* if $\nabla F : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_1$ satisfies

$$\boxed{\begin{aligned} \ln(\nabla F(\lambda x + (1-\lambda)y)) \preceq_{\mathcal{K}_1} \lambda \ln(\nabla F(x)) + (1-\lambda)\ln(\nabla F(y)) \\ \forall x, y \in \mathsf{int}\,\mathcal{K}_1,\;\; \forall\, \lambda \in [0,1] \end{aligned}} \tag{GLC}$$

Recall that $F = f \circ \mathsf{A}$, some examples of $f$ that satisfy (GLC):

# The Class of Gradient Log-Convex Functions

We call $F$ in (P) *gradient log-convex* if $\nabla F : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_1$ satisfies

$$\boxed{\begin{aligned}\ln(\nabla F(\lambda x + (1-\lambda)y)) \preceq_{\mathcal{K}_1} \lambda \ln(\nabla F(x)) + (1-\lambda)\ln(\nabla F(y))\\ \forall x, y \in \mathsf{int}\,\mathcal{K}_1, \;\; \forall\, \lambda \in [0,1]\end{aligned}} \tag{GLC}$$

Recall that $F = f \circ \mathsf{A}$, some examples of $f$ that satisfy (GLC):

- $\mathcal{K}_1 = \mathbb{R}^n_+$, $\mathcal{K}_2$ is any symmetric cone with rank $m$, and

  $$f(y) = m^{-1} \ln\det(y), \quad \forall\, y \in \mathsf{int}\,\mathcal{K}_2$$

  (includes PET and D-OPT).

# The Class of Gradient Log-Convex Functions

We call $F$ in (P) *gradient log-convex* if $\nabla F : \mathsf{int}\,\mathcal{K}_1 \to \mathsf{int}\,\mathcal{K}_1$ satisfies

$$\boxed{\begin{aligned}\ln(\nabla F(\lambda x + (1-\lambda)y)) \preceq_{\mathcal{K}_1} \lambda \ln(\nabla F(x)) + (1-\lambda)\ln(\nabla F(y))\\ \forall x, y \in \mathsf{int}\,\mathcal{K}_1, \;\; \forall\, \lambda \in [0,1]\end{aligned}} \tag{GLC}$$

Recall that $F = f \circ \mathsf{A}$, some examples of $f$ that satisfy (GLC):

- ▷ $\mathcal{K}_1 = \mathbb{R}^n_+$, $\mathcal{K}_2$ is any symmetric cone with rank $m$, and
$$f(y) = m^{-1} \ln\det(y), \quad \forall\, y \in \mathsf{int}\,\mathcal{K}_2$$
(includes PET and D-OPT).
- ▷ $\mathcal{K}_1$ is any *representable* symmetric cone (all except the 27-dimensional exceptional one), $\mathcal{K}_2 = \mathbb{R}^m_+$ and
  - $f(y) = \sum_{j=1}^m w_j \ln y_j$, for all $y > 0$ and $w \in \mathsf{ri}\,\Delta_m$ (includes QST).
  - $f(y) = \ln |||y|||_p := p^{-1} \ln(\sum_{j=1}^m y_j^p)$, for all $y > 0$ and $p \in (0,1]$ (includes RBQP).

# Introducing Euclidean Jordan Algebra (EJA)

- We have developed a Generalized MG (GMG) method for this problem class.

# Introducing Euclidean Jordan Algebra (EJA)

- ▷ We have developed a Generalized MG (GMG) method for this problem class.
- ▷ Our GMG method was developed and analyzed under the framework of EJA.

# Introducing Euclidean Jordan Algebra (EJA)

- We have developed a Generalized MG (GMG) method for this problem class.
- Our GMG method was developed and analyzed under the framework of EJA.
- To understand this method, we will briefly review some basics of EJA.

# Background of EJA

- ▷ We call $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if

# Background of EJA

- ▷ We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space

# Background of EJA

- We call $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$

# Background of EJA

- We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that
    $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.

# Background of EJA

▷ We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if

- $\mathbb{V}$ is a vector space
- $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
- $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.
- Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$.

# Background of EJA

- We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that
    $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.
  - Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$.

- Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

# Background of EJA

- We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.
  - Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$.

- Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

- Any $x \in \mathbb{V}$ has the spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

# Background of EJA

- ▷ We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that
    $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.
  - Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$.

- ▷ Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

- ▷ Any $x \in \mathbb{V}$ has the spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:
  - the eigenvalues $\{\lambda_i(x)\}_{i=1}^n$ are *real*
  - the "eigenvectors" $\{q_i(x)\}_{i=1}^n \subseteq \mathbb{V}$ form a *Jordan frame.*

- ▷ A Jordan frame $\{q_i\}_{i=1}^n \subseteq \mathbb{V}$ satisfy

# Background of EJA

- ▷ We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that
    $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.
  - Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$.

- ▷ Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

- ▷ Any $x \in \mathbb{V}$ has the spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:
  - the eigenvalues $\{\lambda_i(x)\}_{i=1}^n$ are *real*
  - the "eigenvectors" $\{q_i(x)\}_{i=1}^n \subseteq \mathbb{V}$ form a *Jordan frame*.

- ▷ A Jordan frame $\{q_i\}_{i=1}^n \subseteq \mathbb{V}$ satisfy
  - (Completeness) $\sum_{i=1}^n q_i = e$.

# Background of EJA

- We call $(\mathbb{V}, \circ, \langle\cdot,\cdot\rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that
    $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z\rangle = \langle x \circ y, z\rangle$, $\forall\, x, y, z \in \mathbb{V}$.
  - Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$.

- Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

- Any $x \in \mathbb{V}$ has the spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:
  - the eigenvalues $\{\lambda_i(x)\}_{i=1}^n$ are *real*
  - the "eigenvectors" $\{q_i(x)\}_{i=1}^n \subseteq \mathbb{V}$ form a *Jordan frame*.

- A Jordan frame $\{q_i\}_{i=1}^n \subseteq \mathbb{V}$ satisfy
  - (Completeness) $\sum_{i=1}^n q_i = e$.
  - (Orthogonality) $q_i \circ q_j = 0$, $\forall\, i \neq j,\ \ i, j \in [n]$,

# Background of EJA

▷ We call $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if

- $\mathbb{V}$ is a vector space
- $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
- $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z \rangle = \langle x \circ y, z \rangle$, $\forall\, x, y, z \in \mathbb{V}$.
- Let $\| \cdot \|$ be the norm induced by $\langle \cdot, \cdot \rangle$.

▷ Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

▷ Any $x \in \mathbb{V}$ has the spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

- the eigenvalues $\{\lambda_i(x)\}_{i=1}^n$ are *real*
- the "eigenvectors" $\{q_i(x)\}_{i=1}^n \subseteq \mathbb{V}$ form a *Jordan frame*.

▷ A Jordan frame $\{q_i\}_{i=1}^n \subseteq \mathbb{V}$ satisfy

- (Completeness) $\sum_{i=1}^n q_i = e$.
- (Orthogonality) $q_i \circ q_j = 0$, $\forall\, i \neq j,\ \ i, j \in [n]$,
- (Primitiveness and Idempotency) $\|q_i\| = 1$ and $q_i^2 = q_i$, $\forall\, i \in [n]$,

# Background of EJA

- ▷ We call $(\mathbb{V}, \circ, \langle \cdot, \cdot \rangle)$ (or simply $\mathbb{V}$) a EJA with rank $n$ if
  - $\mathbb{V}$ is a vector space
  - $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* operation on $\mathbb{V}$ such that $x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$
  - $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ is *associative*: $\langle x, y \circ z \rangle = \langle x \circ y, z \rangle$, $\forall\, x, y, z \in \mathbb{V}$.
  - Let $\| \cdot \|$ be the norm induced by $\langle \cdot, \cdot \rangle$.

- ▷ Let $e$ be the identity element in $\mathbb{V}$, so that $x \circ e = e \circ x = x$.

- ▷ Any $x \in \mathbb{V}$ has the spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:
  - the eigenvalues $\{\lambda_i(x)\}_{i=1}^n$ are *real*
  - the "eigenvectors" $\{q_i(x)\}_{i=1}^n \subseteq \mathbb{V}$ form a *Jordan frame*.

- ▷ A Jordan frame $\{q_i\}_{i=1}^n \subseteq \mathbb{V}$ satisfy
  - (Completeness) $\sum_{i=1}^n q_i = e$.
  - (Orthogonality) $q_i \circ q_j = 0$, $\forall\, i \neq j,\;\; i, j \in [n]$,
  - (Primitiveness and Idempotency) $\|q_i\| = 1$ and $q_i^2 = q_i$, $\forall\, i \in [n]$,

# Background of EJA

▷ Define $\mathrm{tr}(x) := \sum_{i=1}^{n} \lambda_i(x)$ and $\det(x) := \prod_{i=1}^{n} \lambda_i(x)$.

# Background of EJA

- ▷ Define $\mathrm{tr}(x) := \sum_{i=1}^{n} \lambda_i(x)$ and $\det(x) := \prod_{i=1}^{n} \lambda_i(x)$.
- ▷ W.L.O.G., let $\langle x, y \rangle = \mathrm{tr}(x \circ y)$ so that $\langle e, x \rangle = \mathrm{tr}(x)$.

# Background of EJA

- Define $\mathrm{tr}(x) := \sum_{i=1}^{n} \lambda_i(x)$ and $\det(x) := \prod_{i=1}^{n} \lambda_i(x)$.
- W.L.O.G., let $\langle x, y \rangle = \mathrm{tr}(x \circ y)$ so that $\langle e, x \rangle = \mathrm{tr}(x)$.

Connection between symmetric cones and EJA:

# Background of EJA

- Define $\mathrm{tr}(x) := \sum_{i=1}^{n} \lambda_i(x)$ and $\det(x) := \prod_{i=1}^{n} \lambda_i(x)$.
- W.L.O.G., let $\langle x, y \rangle = \mathrm{tr}(x \circ y)$ so that $\langle e, x \rangle = \mathrm{tr}(x)$.

Connection between symmetric cones and EJA:

> For each symmetric cone $\mathcal{K}$, there exists a unique EJA $\mathbb{V}$ such that $\mathcal{K} \subseteq \mathbb{V}$ and
>
> - $x \in \mathcal{K} \Longleftrightarrow \lambda_1(x), \ldots, \lambda_n(x) \geq 0$
> - $x \in \mathsf{int}\,\mathcal{K} \Longleftrightarrow \lambda_1(x), \ldots, \lambda_n(x) > 0$

# Background of EJA

- Define $\mathrm{tr}(x) := \sum_{i=1}^{n} \lambda_i(x)$ and $\det(x) := \prod_{i=1}^{n} \lambda_i(x)$.
- W.L.O.G., let $\langle x, y \rangle = \mathrm{tr}(x \circ y)$ so that $\langle e, x \rangle = \mathrm{tr}(x)$.

Connection between symmetric cones and EJA:

> For each symmetric cone $\mathcal{K}$, there exists a unique EJA $\mathbb{V}$ such that $\mathcal{K} \subseteq \mathbb{V}$ and
> - $x \in \mathcal{K} \iff \lambda_1(x), \ldots, \lambda_n(x) \geq 0$
> - $x \in \mathsf{int}\,\mathcal{K} \iff \lambda_1(x), \ldots, \lambda_n(x) > 0$
>
> The rank of $\mathcal{K}$ is defined to be the rank of $\mathbb{V}$.

# Generalized MG Method

$$
\boxed{\begin{aligned}
\text{Input:}\quad & x^0 \in \mathsf{ri}\,\mathcal{C} \\
\text{Iterate:}\quad & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\
& x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\
\text{Output:}\quad & \bar{x}^T := (1/T)\textstyle\sum_{t=0}^{T-1} x^t
\end{aligned}} \tag{GMG}
$$

# Generalized MG Method

$$
\begin{array}{|rl|}
\hline
\text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\
\text{Iterate}: & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\
& x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\
\text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t \\
\hline
\end{array}
\qquad \text{(GMG)}
$$

▷ For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

$$\exp(x) = \sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),$$

# Generalized MG Method

$$\begin{array}{rl} \text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\ \text{Iterate}: & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\ & x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\ \text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t \end{array} \qquad \text{(GMG)}$$

- For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

$$\exp(x) = \sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),$$

- If $x^0 \in \mathsf{ri}\,\mathcal{C}$, then $\{x^t\}_{t\geq 0} \subseteq \mathsf{ri}\,\mathcal{C}$.

# Generalized MG Method

$$\begin{array}{|rl|}\hline \text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\ \text{Iterate}: & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\ & x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\ \text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t \\ \hline \end{array} \qquad \text{(GMG)}$$

- For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

$$\exp(x) = \sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),$$

- If $x^0 \in \mathsf{ri}\,\mathcal{C}$, then $\{x^t\}_{t\geq 0} \subseteq \mathsf{ri}\,\mathcal{C}$.
- If $\mathcal{K}_1 = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:

# Generalized MG Method

$$
\begin{array}{rl}
\text{Input :} & x^0 \in \mathsf{ri}\,\mathcal{C} \\
\text{Iterate :} & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\
 & x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\
\text{Output :} & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t
\end{array}
\qquad \text{(GMG)}
$$

- For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

$$
\exp(x) = \sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),
$$

- If $x^0 \in \mathsf{ri}\,\mathcal{C}$, then $\{x^t\}_{t\geq 0} \subseteq \mathsf{ri}\,\mathcal{C}$.
- If $\mathcal{K}_1 = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:
  - both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t)$, $\forall\, i \in [n]$

# Generalized MG Method

$$\boxed{\begin{aligned} \text{Input}: &\quad x^0 \in \mathsf{ri}\,\mathcal{C} \\ \text{Iterate}: &\quad \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\ &\quad x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\ \text{Output}: &\quad \bar{x}^T := (1/T)\textstyle\sum_{t=0}^{T-1} x^t \end{aligned}} \tag{GMG}$$

- For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:
$$\exp(x) = \textstyle\sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),$$
- If $x^0 \in \mathsf{ri}\,\mathcal{C}$, then $\{x^t\}_{t\ge 0} \subseteq \mathsf{ri}\,\mathcal{C}$.
- If $\mathcal{K}_1 = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:
  - both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t)$, $\forall\, i \in [n]$
  - $\mathrm{tr}(\hat{x}^{t+1}) = \langle \nabla F(x^t), x^t \rangle = 1$ (since $F$ is 1-LH)

# Generalized MG Method

$$
\begin{array}{|rl|}
\hline
\text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\
\text{Iterate}: & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\
 & x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\
\text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t \\
\hline
\end{array}
\qquad \text{(GMG)}
$$

- ▷ For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

$$\exp(x) = \sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),$$

- ▷ If $x^0 \in \mathsf{ri}\,\mathcal{C}$, then $\{x^t\}_{t\geq 0} \subseteq \mathsf{ri}\,\mathcal{C}$.
- ▷ If $\mathcal{K}_1 = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:
  - both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t)$, $\forall\, i \in [n]$
  - $\mathrm{tr}(\hat{x}^{t+1}) = \langle \nabla F(x^t), x^t \rangle = 1$ (since $F$ is 1-LH)
  - (GMG) becomes (MG), which only updates eigenvalues

# Generalized MG Method

$$
\begin{array}{rl}
\text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\
\text{Iterate}: & \hat{x}^{t+1} := \exp\{\ln(x^t) + \ln(\nabla F(x^t))\}, \\
& x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\
\text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t
\end{array}
\tag{GMG}
$$

- For $x \in \mathsf{int}\,\mathcal{K}_1$ with spectral decomposition $\sum_{i=1}^n \lambda_i(x) q_i(x)$:

$$\exp(x) = \sum_{i=1}^n \exp(\lambda_i(x)) q_i(x), \quad \ln(x) = \sum_{i=1}^n \ln(\lambda_i(x)) q_i(x),$$

- If $x^0 \in \mathsf{ri}\,\mathcal{C}$, then $\{x^t\}_{t\geq 0} \subseteq \mathsf{ri}\,\mathcal{C}$.

- If $\mathcal{K}_1 = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:
  - both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t), \forall\, i \in [n]$
  - $\mathrm{tr}(\hat{x}^{t+1}) = \langle \nabla F(x^t), x^t \rangle = 1$ (since $F$ is 1-LH)
  - (GMG) becomes (MG), which only updates eigenvalues

- In general, (GMG) updates both eigenvalues and the "eigenvectors", and specializes to all the methods we've seen earlier.

# Convergence Rate of (GMG)

$$
\begin{array}{|rl|}
\hline
\text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\
\text{Iterate}: & \hat{x}^{t+1} := \exp\big(\ln(x^t) + \ln(\nabla F(x^t))\big), \\
& x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\
\text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t \\
\hline
\end{array}
\tag{GMG}
$$

# Convergence Rate of (GMG)

$$\boxed{\begin{aligned} \text{Input}: &\quad x^0 \in \mathsf{ri}\,\mathcal{C} \\ \text{Iterate}: &\quad \hat{x}^{t+1} := \exp\big(\ln(x^t) + \ln(\nabla F(x^t))\big), \\ &\quad x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\ \text{Output}: &\quad \bar{x}^T := (1/T)\textstyle\sum_{t=0}^{T-1} x^t \end{aligned}} \tag{GMG}$$

Theorem (Convergence rate of (GMG))

$$F^* - F(\bar{x}^T) \le \frac{\ln \lambda_{\min}^{-1}(x^0)}{T}, \qquad \forall\, T \ge 1.$$

# Convergence Rate of (GMG)

$$\boxed{\begin{aligned} \text{Input}: &\quad x^0 \in \mathsf{ri}\,\mathcal{C} \\ \text{Iterate}: &\quad \hat{x}^{t+1} := \exp\big(\ln(x^t) + \ln(\nabla F(x^t))\big), \\ &\quad x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\ \text{Output}: &\quad \bar{x}^T := (1/T)\textstyle\sum_{t=0}^{T-1} x^t \end{aligned}} \qquad \text{(GMG)}$$

Theorem (Convergence rate of (GMG))

$$F^* - F(\bar{x}^T) \le \frac{\ln \lambda_{\min}^{-1}(x^0)}{T}, \qquad \forall\, T \ge 1.$$

▷ The convergence rate is data independent — it does not depend on $\mathsf{A}$.

# Convergence Rate of (GMG)

$$\begin{array}{|rl|}\hline \text{Input}: & x^0 \in \mathsf{ri}\,\mathcal{C} \\ \text{Iterate}: & \hat{x}^{t+1} := \exp\big(\ln(x^t) + \ln(\nabla F(x^t))\big), \\ & x^{t+1} := \hat{x}^{t+1}/\mathrm{tr}(\hat{x}^{t+1}). \\ \text{Output}: & \bar{x}^T := (1/T)\sum_{t=0}^{T-1} x^t \\ \hline \end{array} \tag{GMG}$$

**Theorem (Convergence rate of (GMG))**

$$F^* - F(\bar{x}^T) \le \frac{\ln \lambda_{\min}^{-1}(x^0)}{T}, \qquad \forall\, T \ge 1.$$

- ▷ The convergence rate is data independent — it does not depend on $\mathsf{A}$.
- ▷ The optimal choice for the above bound is $x^0 = (1/n)e$, and we have

  $$F^* - F(\bar{x}^T) \le \frac{\ln(n)}{T}, \qquad \forall\, T \ge 1$$

  Recall that $n$ is the rank of $\mathcal{K}_1$.

# Future Work

- Develop other forms of the generalized MG method.
- Discover more applications of (P), particularly when $\mathcal{K}_1$ or $\mathcal{K}_2$ is a Cartesian product of second-order cones.
- Modify the GMG method to accommodate more complicated feasible sets.
- Efficient numerical implementation of GMG method for problems involving matrix variables.

# Thank you!

# References


[Ari72] S. Arimoto. "An algorithm for computing the capacity of arbitrary discrete memoryless channels". In: *IEEE Trans. Inf. Theory* 18.1 (1972), pp. 14–20.

[BBT17] Heinz H. Bauschke, Jérôme Bolte, and Marc Teboulle. "A Descent Lemma Beyond Lipschitz Gradient Continuity: First-Order Methods Revisited and Applications". In: *Math. Oper. Res.* 42.2 (2017), pp. 330–348.

[Coh19] Michael B. Cohen et al. *A near-optimal algorithm for approximating the John Ellipsoid*. arXiv:1905.11580. 2019.

[Cov84] T. Cover. "An algorithm for maximizing expected log investment return". In: *IEEE Trans. Inf. Theory* 30.2 (1984), pp. 369–373.

[Dvu20] Pavel Dvurechensky et al. "Self-Concordant Analysis of Frank-Wolfe Algorithms". In: *Proc. ICML.* 2020, pp. 2814–2824.

[Fed72] V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, 1972.

[FK94] J. Faraut and A. Korányi. *Analysis on Symmetric Cones*. Clarendon Press, 1994.

[Hra04] Zdeněk Hradil et al. "Maximum-Likelihood Methodsin Quantum Mechanics". In: *Quantum State Estimation.* Springer Berlin Heidelberg, 2004, pp. 59–112.

[LCL21] Chien-Ming Lin, Hao-Chung Cheng, and Yen-Huan Li. *Maximum-Likelihood Quantum State Tomography by Cover's Method with Non-Asymptotic Analysis*. arXiv:2110.00747. 2021.

[LFN18] Haihao. Lu, Robert M. Freund, and Yurii. Nesterov. "Relatively Smooth Convex Optimization by First-Order Methods, and Applications". In: *SIAM J. Optim.* 28.1 (2018), pp. 333–354.

[Nes11] Y. Nesterov. "Barrier subgradient method". In: *Math. Program.* (2011), 31—56.

# References

[Nes98] Yu. Nesterov. “Semidefinite relaxation and nonconvex quadratic optimization”. In: *Optim. Methods Softw.* 9.1-3 (1998), pp. 141–160.

[Tod16] Michael J. Todd. *Minimum volume ellipsoids - theory and algorithms*. Vol. 23. SIAM, 2016.

[TWK21] J. Tao, G. Q. Wang, and L. Kong. “The Araki-Lieb-Thirring inequality and the Golden-Thompson inequality in Euclidean Jordan algebras”. In: *Linear Multilinear Algebra* 0.0 (2021), pp. 1–16.

[ZF22] Renbo Zhao and Robert M. Freund. *Analysis of the Frank-Wolfe Method for Convex Composite Optimization involving a Logarithmically-Homogeneous Barrier*. arXiv:2010.08999. 2022.

[Zha21] Renbo Zhao. *Non-Asymptotic Convergence Analysis of the Multiplicative Gradient Algorithm for the Log-Optimal Investment Problems*. arXiv:2109.05601. 2021.

[ZZS13] Ke Zhou, Hongyuan Zha, and Le Song. “Learning Social Infectivity in Sparse Low-rank Networks Using Multi-dimensional Hawkes Processes”. In: *Proc. AISTATS*. 2013, pp. 641–649.