The Generalized Multiplicative Gradient Method for A Class of Convex Symmetric Cone Optimization Problems

Renbo Zhao

Tippie College of Business
University of Iowa

ISMP Montréal
July, 2024

Motivating Example: Positron Emission Tomography

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

Motivating Example: 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].

Motivating Example: 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 $p_j > 0$, $a_j \in \mathbb{R}_+^n$, $a_j \neq 0$ and $\sum_{j=1}^m p_j = 1$.

Motivating Example: 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 $p_j > 0$, $a_j \in \mathbb{R}^n_+$, $a_j \neq 0$ and $\sum_{j=1}^m p_j = 1$.
- $\Delta_n := \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ is the unit simplex in $\mathbb{R}^n$.

Motivating Example: 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 $p_j > 0$, $a_j \in \mathbb{R}^n_+$, $a_j \neq 0$ and $\sum_{j=1}^m p_j = 1$.
- $\Delta_n := \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ is the unit simplex in $\mathbb{R}^n$.
- Multiplicative gradient method: $x^0 \in \mathsf{ri}\,\Delta_n$

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

Motivating Example: 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 $p_j > 0$, $a_j \in \mathbb{R}^n_+$, $a_j \neq 0$ and $\sum_{j=1}^m p_j = 1$.
- $\Delta_n := \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ 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

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

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

The Mystery of MG

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

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

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

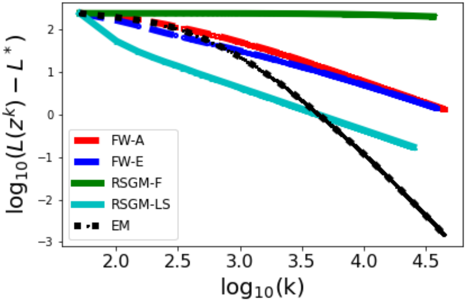
The Mystery of MG

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

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

- 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; ZF20]: 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

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

- 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 [Zha23] 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 [Zha23] 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 [Zha23] 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

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

- ▷ However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- ▷ Zhao [Zha23] 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

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

- ▷ However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- ▷ Zhao [Zha23] 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

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

- ▷ However, MG only has asymptotic convergence guarantees, and the convergence rate has been unknown for about 50 years.
- ▷ Zhao [Zha23] 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.

▷ 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.
- ▷ 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.
- ▷ 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; ZF20] — our generalized MG method contributes to this line of research from a different viewpoint.

D-Optimal Design (D-OPT)

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

D-Optimal Design (D-OPT)

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

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

D-Optimal Design (D-OPT)

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

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

D-Optimal Design (D-OPT)

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

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

D-Optimal Design (D-OPT)

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

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

Quantum State Tomography (QST)

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

Quantum State Tomography (QST)

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

- 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)

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

- 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)

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

- 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)

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

- 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): $q^* := \max_{x\in\{\pm 1\}^n} x^\top A x$ 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): $q^* := \max_{x\in\{\pm1\}^n} x^\top Ax$ 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:

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

  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): $q^* := \max_{x\in\{\pm 1\}^n} x^\top A x$ 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:

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

  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; ZF20]: 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

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

A General Problem Class

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

- ▷ Let Γ be a regular cone in a (finite-dimensional) real vector space $\mathbb{Y}$, and $f : \Gamma \to \mathbb{R} \cup \{-\infty\}$ is a closed and strictly concave function:
  - $\mathsf{int}\,\Gamma \subseteq \mathsf{dom}\, f \subseteq \Gamma \setminus \{0\}$, and $\mathsf{dom}\, f$ invariant under positive scaling.

A General Problem Class

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

▷ Let Γ be a regular cone in a (finite-dimensional) real vector space $\mathbb{Y}$, and $f : \Gamma \to \mathbb{R} \cup \{-\infty\}$ is a closed and strictly concave function:

- $\mathsf{int}\,\Gamma \subseteq \mathsf{dom}\, f \subseteq \Gamma \setminus \{0\}$, and $\mathsf{dom}\, f$ invariant under positive scaling.
- f is C^3 and θ-logarithmically-homogeneous (θ-LH) on $\mathsf{int}\,\Gamma$:

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

A General Problem Class

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

▷ Let Γ be a regular cone in a (finite-dimensional) real vector space $\mathbb{Y}$, and $f : \Gamma \to \mathbb{R} \cup \{-\infty\}$ is a closed and strictly concave function:

- $\mathsf{int}\,\Gamma \subseteq \mathsf{dom}\, f \subseteq \Gamma \setminus \{0\}$, and $\mathsf{dom}\, f$ invariant under positive scaling.
- f is C^3 and θ-logarithmically-homogeneous (θ-LH) on $\mathsf{int}\,\Gamma$:

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

- $\nabla f : \mathsf{int}\,\Gamma \mapsto \mathsf{int}\,\Gamma^*$

A General Problem Class

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

▷ Let Γ be a regular cone in a (finite-dimensional) real vector space $\mathbb{Y}$, and $f : \Gamma \to \mathbb{R} \cup \{-\infty\}$ is a closed and strictly concave function:

- $\mathsf{int}\,\Gamma \subseteq \mathsf{dom}\,f \subseteq \Gamma \setminus \{0\}$, and $\mathsf{dom}\,f$ invariant under positive scaling.
- f is C^3 and θ-logarithmically-homogeneous (θ-LH) on $\mathsf{int}\,\Gamma$:

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

- $\nabla f : \mathsf{int}\,\Gamma \mapsto \mathsf{int}\,\Gamma^*$

▷ $\mathcal{K} \subseteq \mathbb{V}$ is a symmetric cone, where $\mathbb{V}$ is a real inner-product space:

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

A General Problem Class

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

A General Problem Class

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

- For $x \in \mathbb{V}$, $\operatorname{tr}(x)$ denotes the "trace" of x:
  - $\mathbb{V} = \mathbb{R}^n$: $\mathcal{K} = \mathbb{R}^n_+$, $\operatorname{tr}(x) := \sum_{i=1}^n x_i$
  - $\mathbb{V} = \mathbb{S}^n$: $\mathcal{K} = \mathbb{S}^n_+$, $\operatorname{tr}(X) := \sum_{i=1}^n \lambda_i(X)$

A General Problem Class

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

- For $x \in \mathbb{V}$, $\operatorname{tr}(x)$ denotes the "trace" of x:
  - $\mathbb{V} = \mathbb{R}^n$: $\mathcal{K} = \mathbb{R}^n_+$, $\operatorname{tr}(x) := \sum_{i=1}^n x_i$
  - $\mathbb{V} = \mathbb{S}^n$: $\mathcal{K} = \mathbb{S}^n_+$, $\operatorname{tr}(X) := \sum_{i=1}^n \lambda_i(X)$
- $\mathsf{A} : \mathbb{V} \to \mathbb{Y}$ is a linear operator such that

$$\mathsf{A}(\mathsf{int}\,\mathcal{K}) \subseteq \mathsf{int}\,\Gamma \quad \text{and} \quad \mathsf{A}^*(\mathsf{int}\,\Gamma) \subseteq \mathsf{int}\,\mathcal{K}.$$

A General Problem Class

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

▷ For $x \in \mathbb{V}$, $\operatorname{tr}(x)$ denotes the "trace" of x:

- $\mathbb{V} = \mathbb{R}^n$: $\mathcal{K} = \mathbb{R}^n_+$, $\operatorname{tr}(x) := \sum_{i=1}^n x_i$
- $\mathbb{V} = \mathbb{S}^n$: $\mathcal{K} = \mathbb{S}^n_+$, $\operatorname{tr}(X) := \sum_{i=1}^n \lambda_i(X)$

▷ $\mathsf{A} : \mathbb{V} \to \mathbb{Y}$ is a linear operator such that

$$\mathsf{A}(\mathsf{int}\,\mathcal{K}) \subseteq \mathsf{int}\,\Gamma \quad \text{and} \quad \mathsf{A}^*(\mathsf{int}\,\Gamma) \subseteq \mathsf{int}\,\mathcal{K}.$$

▷ Under these conditions, we know that $\nabla F : \mathsf{int}\,\mathcal{K} \to \mathsf{int}\,\mathcal{K}$.

A General Problem Class

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

- ▷ For $x \in \mathbb{V}$, $\operatorname{tr}(x)$ denotes the "trace" of x:
  - $\mathbb{V} = \mathbb{R}^n$: $\mathcal{K} = \mathbb{R}^n_+$, $\operatorname{tr}(x) := \sum_{i=1}^n x_i$
  - $\mathbb{V} = \mathbb{S}^n$: $\mathcal{K} = \mathbb{S}^n_+$, $\operatorname{tr}(X) := \sum_{i=1}^n \lambda_i(X)$
- ▷ $\mathsf{A} : \mathbb{V} \to \mathbb{Y}$ is a linear operator such that

$$\mathsf{A}(\mathsf{int}\,\mathcal{K}) \subseteq \mathsf{int}\,\Gamma \quad \text{and} \quad \mathsf{A}^*(\mathsf{int}\,\Gamma) \subseteq \mathsf{int}\,\mathcal{K}.$$

- ▷ Under these conditions, we know that $\nabla F : \mathsf{int}\,\mathcal{K} \to \mathsf{int}\,\mathcal{K}$.
- ▷ Finally, we require F to be *log-gradient convex* (more details later).

Basics of Euclidean Jordan Algebra (EJA)

▷ We call $(\mathbb{V}, \circ)$ a Jordan algebra (over $\mathbb{R}$) with rank n if

Basics of Euclidean Jordan Algebra (EJA)

▷ We call $(\mathbb{V}, \circ)$ a Jordan algebra (over $\mathbb{R}$) with rank n if

- $\mathbb{V}$ is a real vector space

Basics of Euclidean Jordan Algebra (EJA)

▷ We call $(\mathbb{V}, \circ)$ a Jordan algebra (over $\mathbb{R}$) with rank n if

- $\mathbb{V}$ is a real vector space
- $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* map on $\mathbb{V}$ such that

$$x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$$

Basics of Euclidean Jordan Algebra (EJA)

▷ We call $(\mathbb{V}, \circ)$ a Jordan algebra (over $\mathbb{R}$) with rank n if

- $\mathbb{V}$ is a real vector space
- $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* map on $\mathbb{V}$ such that

$$x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$$

- $\mathbb{V}$ has an identity element e (so that $x \circ e = e \circ x = x$).

▷ Each $x \in \mathbb{V}$ has n eigenvalues $\{\lambda_i(x)\}_{i=1}^n$, and

$$\mathrm{tr}(x) := \sum_{i=1}^n \lambda_i(x) \in \mathbb{R}.$$

Basics of Euclidean Jordan Algebra (EJA)

▷ We call $(\mathbb{V}, \circ)$ a Jordan algebra (over $\mathbb{R}$) with rank n if

- $\mathbb{V}$ is a real vector space
- $\circ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ is a *bilinear* map on $\mathbb{V}$ such that

$$x \circ y = y \circ x, \quad x^2 \circ (x \circ y) = x \circ (x^2 \circ y), \quad \forall\, x, y \in \mathbb{V}.$$

- $\mathbb{V}$ has an identity element e (so that $x \circ e = e \circ x = x$).

▷ Each $x \in \mathbb{V}$ has n eigenvalues $\{\lambda_i(x)\}_{i=1}^n$, and

$$\mathrm{tr}(x) := \textstyle\sum_{i=1}^n \lambda_i(x) \in \mathbb{R}.$$

▷ We call $(\mathbb{V}, \circ)$ *Euclidean* if $\mathrm{tr}(x^2) > 0$ for $x \neq 0$, and we can define an inner product $\langle x, y \rangle := \mathrm{tr}(x \circ y)$ that is *associative*:

$$\langle x \circ y, z \rangle := \langle x, y \circ z \rangle, \quad \forall\, x, y, z \in \mathbb{V}.$$

Spectral Decomposition

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

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

Spectral Decomposition

- ▷ Any x in a EJA $\mathbb{V}$ has a spectral decomposition $x = \sum_{i=1}^{n} \lambda_i(x) q_i(x)$:
  - the eigenvalues $\{\lambda_i(x)\}_{i=1}^{n}$ are *real*
  - $\{q_i(x)\}_{i=1}^{n} \subseteq \mathbb{V}$ is called a *Jordan frame*.
- ▷ A Jordan frame $\{q_i\}_{i=1}^{n} \subseteq \mathbb{V}$ satisfy
  - (Complete) $\sum_{i=1}^{n} q_i = e$
  - (Orthogonal) $q_i \circ q_j = 0, \forall\, i \neq j, \;\; i, j \in [n]$,
  - (Primitive Idempotents) $q_i^2 = q_i$ and $\mathrm{tr}(q_i) = 1, \forall\, i \in [n]$

Spectral Decomposition

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

- ▷ A Jordan frame $\{q_i\}_{i=1}^{n} \subseteq \mathbb{V}$ satisfy
  - (Complete) $\sum_{i=1}^{n} q_i = e$
  - (Orthogonal) $q_i \circ q_j = 0, \forall\, i \neq j, \;\; i, j \in [n]$,
  - (Primitive Idempotents) $q_i^2 = q_i$ and $\mathrm{tr}(q_i) = 1, \forall\, i \in [n]$

- ▷ For any univariate function f, $f(x) := \sum_{i=1}^{n} f(\lambda_i(x)) q_i(x)$ for $x \in \mathbb{V}$.

Spectral Decomposition

- ▷ Any x in a EJA $\mathbb{V}$ has a spectral decomposition $x = \sum_{i=1}^{n} \lambda_i(x) q_i(x)$:
  - the eigenvalues $\{\lambda_i(x)\}_{i=1}^{n}$ are *real*
  - $\{q_i(x)\}_{i=1}^{n} \subseteq \mathbb{V}$ is called a *Jordan frame.*
- ▷ A Jordan frame $\{q_i\}_{i=1}^{n} \subseteq \mathbb{V}$ satisfy
  - (Complete) $\sum_{i=1}^{n} q_i = e$
  - (Orthogonal) $q_i \circ q_j = 0, \forall\, i \neq j,\;\; i, j \in [n]$,
  - (Primitive Idempotents) $q_i^2 = q_i$ and $\mathrm{tr}(q_i) = 1, \forall\, i \in [n]$
- ▷ For any univariate function f, $f(x) := \sum_{i=1}^{n} f(\lambda_i(x)) q_i(x)$ for $x \in \mathbb{V}$.

Connection between symmetric cones and EJA:

Spectral Decomposition

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

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

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

- (Complete) $\sum_{i=1}^{n} q_i = e$
- (Orthogonal) $q_i \circ q_j = 0, \forall\, i \neq j, \;\; i, j \in [n]$,
- (Primitive Idempotents) $q_i^2 = q_i$ and $\mathrm{tr}(q_i) = 1, \forall\, i \in [n]$

▷ For any univariate function f, $f(x) := \sum_{i=1}^{n} f(\lambda_i(x)) q_i(x)$ for $x \in \mathbb{V}$.

Connection between symmetric cones and EJA:

For each symmetric cone $\mathcal{K}$, there exists a unique EJA $\mathbb{V}$ such that $\mathcal{K} := \{x^2 : x \in \mathbb{V}\}$ and

▷ $x \in \mathcal{K} \Longleftrightarrow \lambda_1(x), \ldots, \lambda_n(x) \geq 0$

▷ $x \in \mathsf{int}\, \mathcal{K} \Longleftrightarrow \lambda_1(x), \ldots, \lambda_n(x) > 0$

Generalized MG Method

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

Generalized MG Method

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

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

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

Generalized MG Method

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

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

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

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

Generalized MG Method

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

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

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

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

Generalized MG Method

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

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

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

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

- ▷ If $\mathcal{K} = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:
  - both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t)/\theta$, $\forall\, i \in [n]$

Generalized MG Method

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

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

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

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

Generalized MG Method

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

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

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

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

▷ If $\mathcal{K} = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:

- both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t)/\theta$, $\forall\, i \in [n]$
- $\mathrm{tr}(\hat{x}^{t+1}) = \langle \nabla F(x^t), x^t \rangle / \theta = 1$ (since F is 1-LH)
- (GMG) becomes (MG), which only updates eigenvalues

Generalized MG Method

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

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

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

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

▷ If $\mathcal{K} = \mathbb{R}^n_+$, then $x = \sum_{i=1}^n x_i e_i$:

- both $\exp(\cdot)$ and $\ln(\cdot)$ are element-wise $\Longrightarrow \hat{x}^{t+1} = x^t \circ \nabla F(x^t)/\theta$, $\forall\, i \in [n]$
- $\mathrm{tr}(\hat{x}^{t+1}) = \langle \nabla F(x^t), x^t \rangle/\theta = 1$ (since F is 1-LH)
- (GMG) becomes (MG), which only updates eigenvalues

▷ In general, (GMG) updates both eigenvalues and the Jordan frame, and specializes to all the methods we've seen earlier.

Convergence Rate of (GMG)

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

Convergence Rate of (GMG)

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

Theorem (Convergence rate of (GMG))

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

Convergence Rate of (GMG)

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

Theorem (Convergence rate of (GMG))

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

- ▷ The convergence rate is data independent — it does not depend on A.

Convergence Rate of (GMG)

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

Theorem (Convergence rate of (GMG))

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

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

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

(Recall that n is the rank of the EJA associated with $\mathcal{K}$.)

The Class of Gradient Log-Convex Functions

Let $\mathrm{PI}(\mathbb{V}) := \{x \in \mathbb{V} : x^2 = x, \mathrm{tr}(x) = 1\}$ be the primitive idempotents in $\mathbb{V}$.

We call $F : \mathcal{K} \to \mathbb{R} \cup \{-\infty\}$ in (P) *log-gradient convex* if

$$\forall u \in \mathrm{PI}(\mathbb{V}), \quad x \mapsto \langle \ln \nabla F(x), u \rangle \quad \text{is convex on } \mathsf{int}\,\mathcal{K} \tag{GLC}$$

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

The Class of Gradient Log-Convex Functions

Let $\mathrm{PI}(\mathbb{V}) := \{x \in \mathbb{V} : x^2 = x, \mathrm{tr}(x) = 1\}$ be the primitive idempotents in $\mathbb{V}$.

We call $F : \mathcal{K} \to \mathbb{R} \cup \{-\infty\}$ in (P) *log-gradient convex* if

$$\forall u \in \mathrm{PI}(\mathbb{V}), \quad x \mapsto \langle \ln \nabla F(x), u \rangle \quad \text{is convex on } \mathsf{int}\, \mathcal{K} \tag{GLC}$$

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

- $\mathcal{K} = \mathbb{R}^n_+$ and Γ is the closure of the hyperbolicity cone associated with a *complete* hyperbolic polynomial $p : \mathbb{Y} \to \mathbb{R}$, then

  $$f(y) = \ln p(y), \quad \forall\, y \in \mathsf{int}\, \Gamma$$

  (includes PET and D-OPT).

The Class of Gradient Log-Convex Functions

Let $\mathrm{PI}(\mathbb{V}) := \{x \in \mathbb{V} : x^2 = x, \mathrm{tr}(x) = 1\}$ be the primitive idempotents in $\mathbb{V}$.

We call $F : \mathcal{K} \to \mathbb{R} \cup \{-\infty\}$ in (P) *log-gradient convex* if

$$\forall u \in \mathrm{PI}(\mathbb{V}), \quad x \mapsto \langle \ln \nabla F(x), u \rangle \quad \text{is convex on } \mathsf{int}\,\mathcal{K} \tag{GLC}$$

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

- $\mathcal{K} = \mathbb{R}^n_+$ and Γ is the closure of the hyperbolicity cone associated with a *complete* hyperbolic polynomial $p : \mathbb{Y} \to \mathbb{R}$, then

  $$f(y) = \ln p(y), \quad \forall\, y \in \mathsf{int}\,\Gamma$$

  (includes PET and D-OPT).

- $\mathcal{K}$ is any *representable* symmetric cone and $\Gamma = \mathbb{R}^m_+$, then
  - $f(y) = \sum_{j=1}^m w_j \ln y_j$, for all $y > 0$ and $w \in \mathsf{ri}\,\Delta_m$ (includes QST).
  - $f(y) = \ln \left\lvert\!\left\lvert\!\left\lvert y \right\rvert\!\right\rvert\!\right\rvert_p$, for all $y > 0$ and $p \in (0, 1]$ (includes RBQP).

Thank you!

Proof Sketch

Lemma 1 (An important conic inequality)

For all $t \geq 0$*, we have* $\mathrm{tr}(\hat{x}^{t+1}) \leq 1$ *and hence*

$$\ln(x^{t+1}) \succeq_{\mathcal{K}_1} \ln(x^t) + \ln(\nabla F(x^t)).$$

Proof Sketch

Lemma 1 (An important conic inequality)

For all $t \geq 0$, *we have* $\mathrm{tr}(\hat{x}^{t+1}) \leq 1$ *and hence*

$$\ln(x^{t+1}) \succeq_{\mathcal{K}_1} \ln(x^t) + \ln(\nabla F(x^t)).$$

This inequality essentially follows from the Golden-Thompson inequality for EJA in Tao et al. [TWK21].

To telescope, need the assumption that $\ln \nabla F(\cdot)$ is convex on $\mathsf{ri}\, \mathcal{C}$ (w.r.t. $\mathcal{K}_1$).

Lemma 2 (A growth bound of F)

For any $x \in \mathsf{ri}\, \mathcal{C}$ *and any optimal solution* x^*, $F^* - F(x) \leq \ln(\langle \nabla F(x), x^* \rangle)$.

Proof Sketch

Lemma 1 (An important conic inequality)

For all $t \geq 0$, *we have* $\mathrm{tr}(\hat{x}^{t+1}) \leq 1$ *and hence*

$$\ln(x^{t+1}) \succeq_{\mathcal{K}_1} \ln(x^t) + \ln(\nabla F(x^t)).$$

This inequality essentially follows from the Golden-Thompson inequality for EJA in Tao et al. [TWK21].

To telescope, need the assumption that $\ln \nabla F(\cdot)$ is convex on $\mathsf{ri}\,\mathcal{C}$ (w.r.t. $\mathcal{K}_1$).

Lemma 2 (A growth bound of F)

For any $x \in \mathsf{ri}\,\mathcal{C}$ *and any optimal solution* x^*, $F^* - F(x) \leq \ln(\langle \nabla F(x), x^* \rangle)$.

This bound is solely due to the 1-log-homogeneity of F.

Proof Sketch

Lemma 1 (An important conic inequality)

For all $t \geq 0$, *we have* $\operatorname{tr}(\hat{x}^{t+1}) \leq 1$ *and hence*

$$\ln(x^{t+1}) \succeq_{\mathcal{K}_1} \ln(x^t) + \ln(\nabla F(x^t)).$$

This inequality essentially follows from the Golden-Thompson inequality for EJA in Tao et al. [TWK21].

To telescope, need the assumption that $\ln \nabla F(\cdot)$ is convex on $\mathsf{ri}\,\mathcal{C}$ (w.r.t. $\mathcal{K}_1$).

Lemma 2 (A growth bound of F)

For any $x \in \mathsf{ri}\,\mathcal{C}$ *and any optimal solution* x^*, $F^* - F(x) \leq \ln(\langle \nabla F(x), x^* \rangle)$.

This bound is solely due to the 1-log-homogeneity of F.

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.

[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.

[ZF20] Renbo Zhao and Robert M. Freund. *Analysis of the Frank-Wolfe Method for Convex Composite Optimization involving a Logarithmically-Homogeneous Barrier*. arXiv:2010.08999. 2020.

[Zha23] Renbo Zhao. "Convergence Rate Analysis of the Multiplicative Gradient Algorithm for PET-Type Problems". In: *Oper. Res. Lett.* 51.1 (2023), 26 — 32.

[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.