Optimization Techniques for Alphabet-Constrained Signal Design

Mojtaba Soltanalian

Department of Electrical Engineering California Institute of Technology

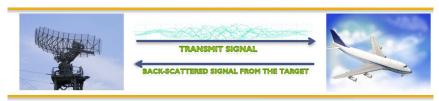
Stanford EE- ISL Mar. 2015

Outline

- Signal Design: what is this all about?
- 2 Alternating Projections on Converging Sets (ALPS-CS)
- Power Method-Like Iterations
- 4 MERIT

Signal Design- some applications

- Signal design for active sensing.
 Goal: To acquire (or preserve) the maximum information from the desirable sources in the environment.
- Signal is a medium to collect information.

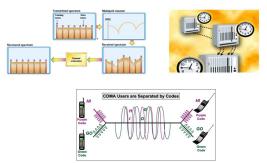


 The research in this area is focused on the design and optimization of probing signals to improve target detection performance, as well as the target location and speed estimation accuracy.

Signal Design- some applications

Signal design for communications.

Goal: To <u>transfer</u> the maximum information among chosen agents in the network.



• Applications in *Channel Estimation, Code-Division Multiple-Access* (CDMA) Schemes, Synchronization, Beamforming, . . .

Signal Design- some applications

Signal design for life sciences.

Goal: To make the <u>best identification</u> of the living organism, usually by *maximal excitation*.







Signal Design- Keywords

- Waveform design and diversity (signal processing- communications)
- Input design (control- system identification)
- Sequence design (signal processing- information theorycommunications- mathematics)
- Stimulus design, excitation design.

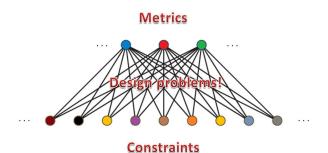
Signal Design- Metrics

- Mean-Square Error (MSE) of parameter estimation
- Signal-to-Noise Ratio (SNR) of the received data
- Information-Theoretic criteria
- Auto/Cross Correlation Sidelobe metrics
- Excitation metrics
- Stability metrics
- Secrecy metrics
- . . .

Signal Design- Constraints

- Energy
- Peak-to-Average Power Ratio (PAPR, PAR)
- Unimodularity (being Constant-Modulus)
- Finite or Discrete-Alphabet (integer, binary, m-ary constellation)
- . . .

Signal Design



Many of these problem are shown to be NP-hard; Many others are *deemed* to be difficult!

Challenges:

How to handle signal constraints?

-and how to do it fast?

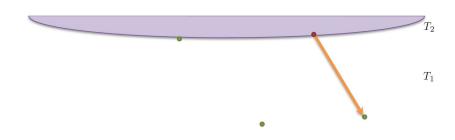
Signal Design- Methodologies

Useful design techniques:

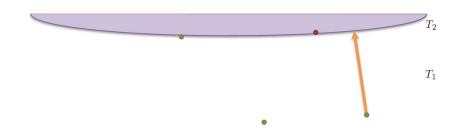
- Alternating Projections on Converging Sets (ALPS-CS)
- Power Method-Like Iterations
- MERIT

- Alternating Projections for signal design
- Alternating Projections convex vs non-convex, finite-alphabet

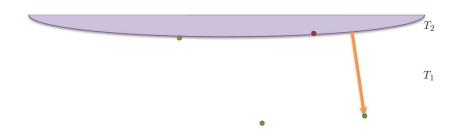
- Alternating Projections convex vs non-convex, finite-alphabet
- Example: T_1 a set with 3 elements (green dots); T_2 a convex set.



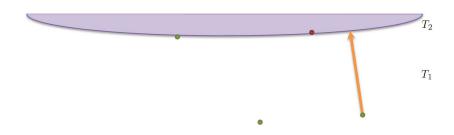
- Alternating Projections convex vs non-convex, finite-alphabet
- Example: T_1 a set with 3 elements (green dots); T_2 a convex set.



- Alternating Projections convex vs non-convex, finite-alphabet
- Example: T_1 a set with 3 elements (green dots); T_2 a convex set.



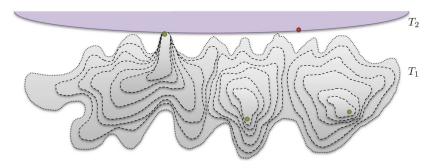
- Alternating Projections convex vs non-convex, finite-alphabet
- Example: T_1 a set with 3 elements (green dots); T_2 a convex set.



Significant possibility of getting stuck in a poor "solution".

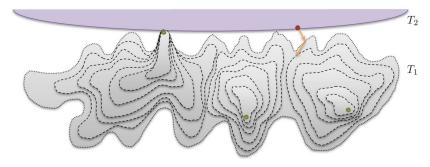
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



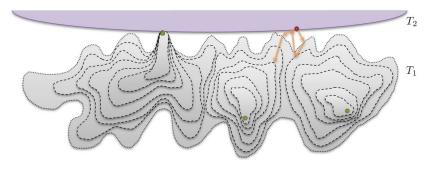
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



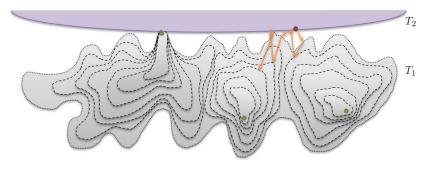
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



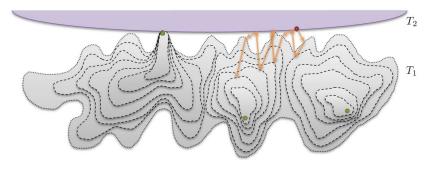
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



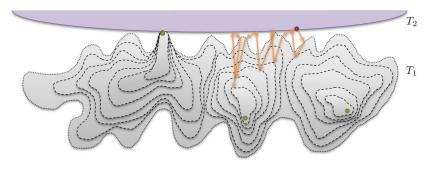
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



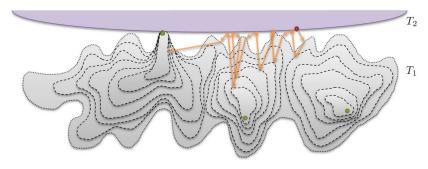
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



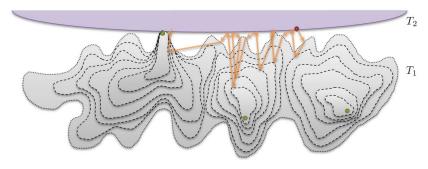
Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.

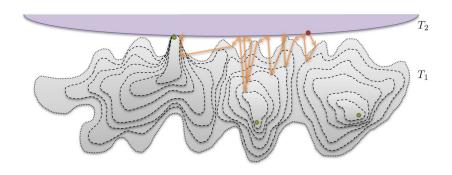


Central Idea:

To replace the "tricky" set with a well-behaved (perhaps compact/convex) set that in limit converges to the "tricky" set of interest! Then we employ the typical alternating projections, while the replaced set, at each iteration, gets closer to the "tricky" set.



Why should this work?



 Selection of the converging sets can be done by choosing a converging function. Example ($\nu > 0$)

(a)
$$T = \mathbb{R} - \{0\}, \ T^{\dagger} = \{-1, 1\}:$$

$$f(t, s) = \operatorname{sgn}(t) \cdot |t|^{e^{-\nu s}}; \tag{1}$$

(b)
$$T = \mathbb{C} - \{0\}, \ T^{\dagger} = \{\zeta \in \mathbb{C} \mid |\zeta| = 1\}:$$

$$f(t,s) = |t|^{e^{-\nu s}} \cdot e^{j \arg(t)}. \tag{2}$$

- If the associated function f is **monotonic** and **identity**, we can show the convergence.
- How to choose f "optimally"? (open problem)
- For more details, see

"Computational Design of Sequences with Good Correlation Properties," IEEE Transactions on Signal Processing, vol. 60, no. 5, pp. 2180-2193, 2012.

A Numerical Example

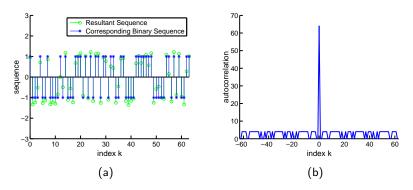


Figure: Design of a binary sequence of length 64 with good periodic auto-correlation using ALPS-CS. (a) the sequence provided by ALPS-CS when stopped, along with the corresponding binary sequence (obtained by clipping). The autocorrelation of the binary sequence is shown in (b).

ALPS-CS requires a design of the alternating projections as well as a suitable choice of converging function.

Let's see a simpler method!

Many signal design problems can be formulated as (a sequence of)
quadratic programs (QPs): SNR maximization, CRLB minimization,
MSE minimization, beam-pattern matching, optimization of
information-theoretic criteria, low-rank recovery, maximum-likelihood.

- Many signal design problems can be formulated as (a sequence of) quadratic programs (QPs): SNR maximization, CRLB minimization, MSE minimization, beam-pattern matching, optimization of information-theoretic criteria, low-rank recovery, maximum-likelihood.
- Some may need more sophisticated ideas for transformation to QP: fractional programming, MM algorithm, cyclic optimization, over-parametrization, etc.

Formulation:

$$\max_{\boldsymbol{s} \in \mathbb{C}^n}. \qquad \boldsymbol{s}^H \boldsymbol{R} \boldsymbol{s}$$
s. t.
$$\boldsymbol{s} \in \Omega$$

$$(\Omega : \text{search space})$$

We can usually assume that the signal power is fixed: (why?)

$$\max_{\mathbf{s} \in \mathbb{C}^n}. \quad \mathbf{s}^H \mathbf{R} \mathbf{s}$$
s. t. $\mathbf{s} \in \Omega$

$$\|\mathbf{s}\|_2^2 = n.$$

Central Idea

• Assume **R** is positive definite (or make it so).

Central Idea

- Assume **R** is positive definite (or make it so).
- Start from some feasible $s = s^{(0)}$, and form the sequence:

$$s^{(t+1)} = \operatorname{Proj}_{\Omega}\left(\mathbf{R}s^{(t)}\right)$$
 (5)

where $ext{Proj}_{\Omega}(\pmb{x}) = ext{arg min}_{\pmb{s} \in \Omega, \, \|\pmb{s}\|_2^2 = n} \, \|\pmb{s} - \pmb{x}\|_2$ denotes the nearest vector in the search space (l_2 -norm sense).

Central Idea

- Assume **R** is positive definite (or make it so).
- Start from some feasible $s = s^{(0)}$, and form the sequence:

$$s^{(t+1)} = \operatorname{Proj}_{\Omega}\left(Rs^{(t)}\right)$$
 (5)

where $\operatorname{Proj}_{\Omega}(x) = \arg\min_{s \in \Omega, \|s\|_{2}^{2} = n} \|s - x\|_{2}$ denotes the nearest vector in the search space (l_2 -norm sense).

 The above power method-like iterations lead to a monotonic increase of the QP objective. \rightarrow convergence!

Central Idea

- Assume **R** is positive definite (or make it so).
- Start from some feasible $s = s^{(0)}$, and form the sequence:

$$s^{(t+1)} = \operatorname{Proj}_{\Omega}\left(Rs^{(t)}\right)$$
 (5)

where $\operatorname{Proj}_{\Omega}(x) = \arg\min_{s \in \Omega, \|s\|_{2}^{2} = n} \|s - x\|_{2}$ denotes the nearest vector in the search space (l_2 -norm sense).

- The above power method-like iterations lead to a monotonic increase of the QP objective. \rightarrow convergence!
- → This is very fast! (No matrix inversion needed.)

Let's see some examples— Constraints:

• Unimodular s ($\Omega = \{s : |s| = 1\}^n$):

$$\mathbf{s}^{(t+1)} = \exp\left(j\arg\left(\mathbf{R}\mathbf{s}^{(t)}\right)\right)$$
. . . just keep the phase.

Let's see some examples— Constraints:

• Unimodular s ($\Omega = \{s : |s| = 1\}^n$):

$$\mathbf{s}^{(t+1)} = \exp\left(j\arg\left(\mathbf{R}\mathbf{s}^{(t)}\right)\right)$$
. . . just keep the phase.

• Binary s ($\Omega = \{-1, +1\}^n$):

$$\mathbf{s}^{(t+1)} = \operatorname{sgn}\left(\Re\left(\mathbf{R}\mathbf{s}^{(t)}\right)\right)$$
. . . just keep the sign.

Let's see some examples— Constraints:

• Unimodular s ($\Omega = \{s : |s| = 1\}^n$):

$$s^{(t+1)} = \exp\left(j\arg\left(Rs^{(t)}\right)\right)$$
. . . just keep the phase.

• Binary s ($\Omega = \{-1, +1\}^n$):

$$s^{(t+1)} = \operatorname{sgn}\left(\Re\left(\mathbf{R}s^{(t)}\right)\right)$$
... just keep the sign.

• Sparse s ($||s||_0 \le k$):

. . . just keep the k largest values of $Rs^{(t)}$ (and scale). (8)

Transformations to QP

Example (beam-pattern matching, low-coherence sensing for radar): Given positive-definite $\{R_k\}_{k=1}^t$ and non-negative $\{d_k\}_{k=1}^t$,

$$\min_{\boldsymbol{s} \in \mathbb{C}^n} . \quad \sum_{k=1}^t |\boldsymbol{s}^H \boldsymbol{R}_k \boldsymbol{s} - d_k|^2
s. t. \qquad \boldsymbol{s} \in \Omega
\|\boldsymbol{s}\|_2^2 = n.$$
(9)

Transformations to QP

Example (beam-pattern matching, low-coherence sensing for radar): Given positive-definite $\{R_k\}_{k=1}^t$ and non-negative $\{d_k\}_{k=1}^t$,

$$\min_{\mathbf{s} \in \mathbb{C}^n} . \quad \sum_{k=1}^t |\mathbf{s}^H \mathbf{R}_k \mathbf{s} - d_k|^2
\text{s. t.} \qquad \mathbf{s} \in \Omega
\|\mathbf{s}\|_2^2 = n.$$
(9)

Over-parametrized "almost-equivalent" form:

$$\begin{aligned} & \min_{\boldsymbol{s}, \{\boldsymbol{u}_k\}}. \quad \sum_{k=1}^t \left\| \boldsymbol{R}_k^{1/2} \boldsymbol{s} - \sqrt{d_k} \boldsymbol{u}_k \right\|^2 \\ & \text{s. t.} \qquad \boldsymbol{s} \in \Omega, \ \|\boldsymbol{s}\|_2^2 = n; \\ & \|\boldsymbol{u}_k\|_2 = 1, \ 1 < k < t. \end{aligned}$$

(10)

Transformations to QP

Example (beam-pattern matching, low-coherence sensing for radar):

• Over-parametrized "almost-equivalent" form:

$$\begin{aligned} \min_{\boldsymbol{s}, \{\boldsymbol{u}_k\}} \cdot \quad & \sum_{k=1}^t \left\| \boldsymbol{R}_k^{1/2} \boldsymbol{s} - \sqrt{d_k} \boldsymbol{u}_k \right\|^2 \\ \text{s. t.} \quad & \boldsymbol{s} \in \Omega, \ \|\boldsymbol{s}\|_2^2 = n; \\ & \|\boldsymbol{u}_k\|_2 = 1, \ 1 \le k \le t. \end{aligned}$$

ullet Minimization with respect to $oldsymbol{s}$ boils down to

$$\begin{aligned} & \min_{\boldsymbol{s} \in \mathbb{C}^n}. & \left(\begin{array}{c} \boldsymbol{s} \\ 1 \end{array}\right)^H \left(\begin{array}{cc} \sum_{k=1}^t \boldsymbol{R}_k & \sum_{k=1}^t \sqrt{d_k} \boldsymbol{R}_k^{1/2} \boldsymbol{u}_k \\ \sum_{k=1}^t \sqrt{d_k} \boldsymbol{u}_k^H \boldsymbol{R}_k^{1/2} & 0 \end{array}\right) \left(\begin{array}{c} \boldsymbol{s} \\ 1 \end{array}\right) \\ & \text{s. t.} & \boldsymbol{s} \in \Omega \\ & \|\boldsymbol{s}\|_2^2 = n. \end{aligned}$$

Transformations to QP

For other examples, see

- Information-theoretic metrics:
 - * "Unified Optimization Framework for Multi-Static Radar Code Design Using Information-Theoretic Criteria," IEEE Transactions on Signal Processing, vol. 61, no. 21, pp. 5401-5416, 2013.
- MSE:
 - * "Optimized Transmission for Centralized Estimation in Wireless Sensor Networks," Preprint.
 - * "Training Signal Design for Massive MIMO Channel Estimation," Preprint.

- For more details about power method-like iterations, see
 - * "Designing Unimodular Codes Via Quadratic Optimization." IEEE Transactions on Signal Processing, vol. 62, no. 5, pp. 1221-1234, 2014.
 - * "Joint Design of the Receive Filter and Transmit Sequence for Active Sensing," IEEE Signal Processing Letters, vol. 20, no. 5, pp. 423-426, 2013.

Power method is fast, but doesn't reveal any information on where the signal quality stands with regard to the optimal value of the design problem . . .

MERIT

- MERIT stands for a Monotonically ERror-Bound Improving Technique for Mathematical Optimization.
- It's a computational framework to obtain sub-optimality guarantees along with the approximate solutions.
- You want to know how much the solution can be trusted . . .

The Central Idea

Let $\mathcal{P}(v,x)$ be an optimization problem *structure* with given and optimization variables partitioned as (v,x).

Example

$$\mathbf{X} = \operatorname{arg\ max} \quad \operatorname{tr}(\mathbf{RX})$$
 variable partitioning $\mathbf{R}, \mathbf{Q}, t \to v$ s.t. $\operatorname{tr}(\mathbf{QX}) \le t$ \Longrightarrow $\mathbf{X} \to x$

The Central Idea

Now suppose $\mathcal{P}(v,x)$ is a "difficult" optimization problem; however,

- A sequence v_1, v_2, v_3, \cdots of v can be constructed such that the associated global optima of the problem, viz. $x_k = \arg\max_x \mathcal{P}(v_k, x)$ are known for any v_k , and the "distance" between v and v_k , is decreasing with k;
- A sub-optimality guarantee of the obtained solutions x_k can be efficiently computed using the distance between v and v_k .

The Central Idea

Then, computational sub-optimality guarantees is obtained along with the approximate solutions, that might

- outperform existing analytically derived sub-optimality guarantees, or
- be the only class of sub-optimality guarantees in cases where no a priori known analytical guarantees are available for the given problem.

An example:

Unimodular Quadratic Programming (UQP)

$$UQP: \max_{s \in \Omega^n} s^H Rs$$
 (11)

where $\mathbf{R} \in \mathbb{C}^{n \times n}$ is a given Hermitian matrix, and Ω represents the unit circle, i.e. $\Omega = \{s \in \mathbb{C} : |s| = 1\}$.

UQP is NP-hard.

UQP: $\max_{s \in \Omega^n} s^H R s$

MERIT:

Build a sequence of matrices (for which the UQP global optima are known) whose distance from the given matrix R is decreasing.

UQP: $\max s^H R s$

Theorem

Let $\mathcal{K}(\mathbf{s})$ represent the set of matrices **R** for which a given $\mathbf{s} \in \Omega^n$ is the global optimizer of UQP. Then

- **1** $\mathcal{K}(s)$ is a convex cone.
- 2 For any two vectors $\mathbf{s}_1, \mathbf{s}_2 \in \Omega^n$, the one-to-one mapping (where $s_0 = s_1^* \odot s_2$

$$\mathbf{R} \in \mathcal{K}(\mathbf{s}_1) \Longleftrightarrow \mathbf{R} \odot (\mathbf{s}_0 \mathbf{s}_0^H) \in \mathcal{K}(\mathbf{s}_2)$$
 (12)

holds among the matrices in $\mathcal{K}(\mathbf{s}_1)$ and $\mathcal{K}(\mathbf{s}_2)$.



UQP: Approximation of $\mathcal{K}(s)$

Theorem

For any given $\mathbf{s} = (e^{j\phi_1}, \cdots, e^{j\phi_n})^T \in \Omega^n$, let $\mathcal{C}(\mathbf{V}_s)$ represent the convex cone of matrices $V_s = D \odot (ss^H)$ where D is any real-valued symmetric matrix with non-negative off-diagonal entries. Also let C_s represent the convex cone of matrices with s being their dominant eigenvector (i.e the eigenvector corresponding to the maximal eigenvalue). Then for any $\mathbf{R} \in \mathcal{K}(\mathbf{s})$, there exists $\alpha_0 \geq 0$ such that for all $\alpha \geq \alpha_0$,

$$\mathbf{R} + \alpha \mathbf{s} \mathbf{s}^H \in \mathcal{C}(\mathbf{V}_{\mathbf{s}}) \oplus \mathcal{C}_{\mathbf{s}}$$
 (13)

where \oplus stands for the Minkowski sum of the two sets.

UQP: Approximation of $\mathcal{K}(s)$

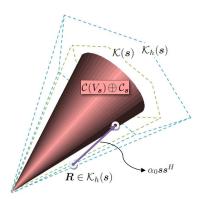


Figure: An illustration of the cone approximation technique used for MERIT's derivation in unimodular quadratic programming.

UQP: MERIT Objective

• Using the previous results, we build a sequence of matrices (for which the UQP global optima are known) whose distance from the given matrix \mathbf{R} is decreasing.

UQP: MERIT Objective

- Using the previous results, we build a sequence of matrices (for which the UQP global optima are known) whose distance from the given matrix \mathbf{R} is decreasing.
- Instead of the original UQP, we consider the optimization problem:

$$\min_{\boldsymbol{s} \in \Omega^n, \boldsymbol{Q}_1 \in \mathcal{C}_1, \boldsymbol{P}_1 \in \mathcal{C}(\boldsymbol{V}_1)} \| \boldsymbol{R} - (\boldsymbol{Q}_1 + \boldsymbol{P}_1) \odot (\boldsymbol{s} \boldsymbol{s}^H) \|_F$$
 (14)

UQP: MERIT Objective

- Using the previous results, we build a sequence of matrices (for which the UQP global optima are known) whose distance from the given matrix R is decreasing.
- Instead of the original UQP, we consider the optimization problem:

$$\min_{\boldsymbol{s} \in \Omega^n, \boldsymbol{Q}_1 \in \mathcal{C}_1, \boldsymbol{P}_1 \in \mathcal{C}(\boldsymbol{V}_1)} \| \boldsymbol{R} - (\boldsymbol{Q}_1 + \boldsymbol{P}_1) \odot (\boldsymbol{s} \boldsymbol{s}^H) \|_F$$
 (14)

• $(Q_1 + P_1) \odot (ss^H)$ will get close to R.

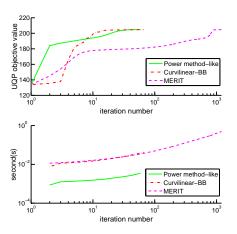


Figure: A comparison of power method-like iterations, the curvilinear search with Barzilai-Borwein step size, and MERIT: (top) the UQP objective; (bottom) the required time for approximating UQP solution (n=10) with same initialization.

n	Rank (d)	$\#$ problems for which $\gamma=1$	Average γ	Average SDR time Average MERIT time
8	2	17	0.9841	1.08
	8	16	0.9912	0.81
	2	15	0.9789	2.08
16	4	13	0.9773	0.95
	16	4	0.9610	0.92

Table: Comparison of the performance of MERIT and SDR when solving UQP for 20 random positive definite matrices of different sizes n and ranks d.

MERIT

- For more details on MERIT, see
 - * "Designing Unimodular Codes Via Quadratic Optimization," IEEE Transactions on Signal Processing, vol. 62, no. 5, pp. 1221-1234, 2014.
 - * "Beyond Semidefinite Relaxtion: Basis Banks and Computationally Enhanced Guarantees," Submitted to IEEE International Symposium on Information Theory (ISIT), Hong Kong, 2015.

Summary- what we discussed?

- Various signal design problems arise in practice.
- Signal design methodologies:
 - Alternating Projections on Converging Sets (ALPS-CS)
 - Power Method-Like Iterations
 - MERIT