ELEG5481

SIGNAL PROCESSING OPTIMIZATION TECHNIQUES

0. INTRODUCTION

A General Formulation of Optimization

$$\begin{aligned} &\min\ f_0(x) \\ &\text{s.t. } f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{aligned}$$

$x = [\ x_1, \ldots, x_n\ ]^T \in \mathbf{R}^n$ optimization variables

$f_0 : \mathbf{R}^n \to \mathbf{R}$ objective function

$f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$ constraint functions

Our aim is to find an **optimal solution** $x^\star$ that minimizes f_0 whilst satisfying all the constraints.

Why does optimization concern me?

- Optimization has found applications in a wide variety of areas such as finance, statistics, and engineering, to name a few.
- In engineering, it plays a key role in solving or handling numerous (and sometimes very hard) problems in control, circuit design, networks, signal processing, and communications.
- It is an important enough topic that we should know at least a bit about it.

Example: Diet Problem

- x_i is the quantity of food i.
- Each unit of food i has a cost of c_i.
- One unit of food j contains an amount a_{ij} of nutrient i.
- We want nutrient i to be at least equal to b_i.
- **Problem:** find the cheapest diet such that the minimum nutrient requirements are fulfilled.

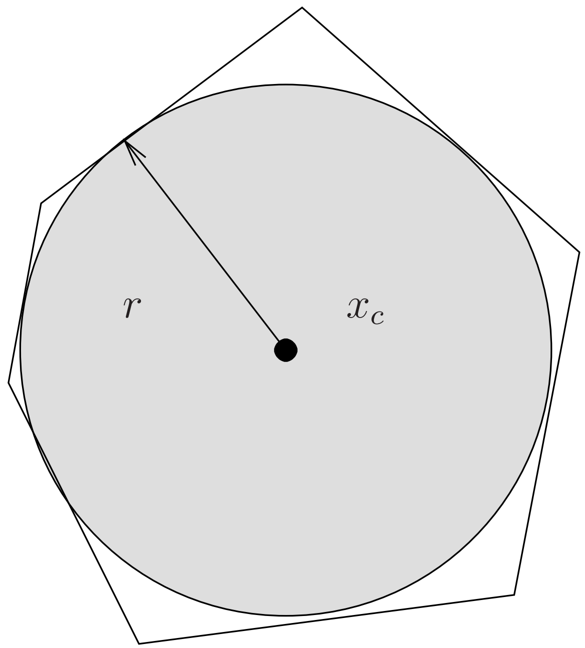
This problem can be formulated as:

$$\begin{aligned} \min \ & \sum_{i=1}^{n} c_i x_i \\ \text{s.t.} \ & \sum_{j=1}^{n} a_{ij} x_j \geq b_i, \quad i = 1, 2, \ldots, m \\ & x_i \geq 0, \quad i = 1, 2, \ldots, n \end{aligned}$$

This is a **linear program**.

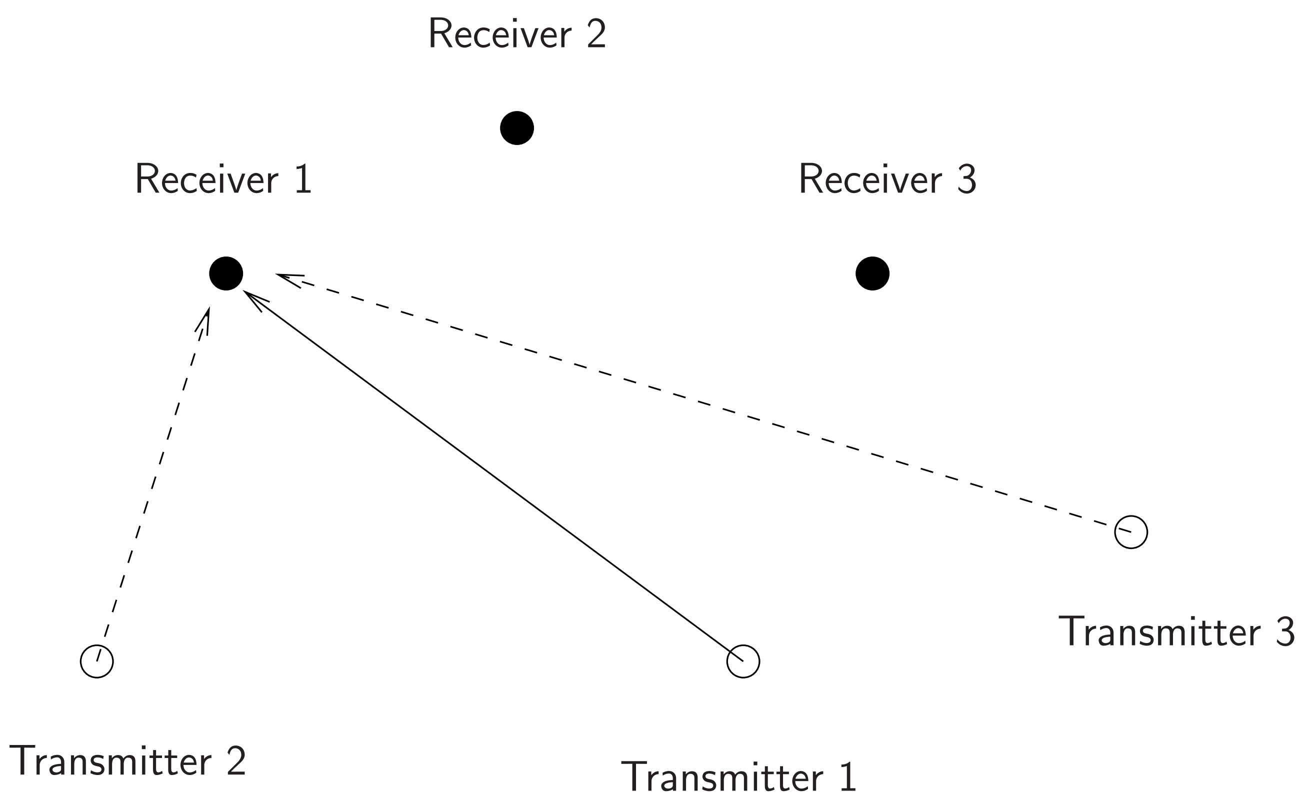
Example: Chebychev Center

- Let a norm ball $B(x_c, r) = \{ \, x \mid \|x_c - x\|_2 \leq r \, \}$, & a polyhedron $\mathcal{P} = \{ \, x \mid a_i^T x \leq b_i, \; i = 1, \ldots, m \, \}$.
- **Problem:** Find the largest ball inside a polyhedron $\mathcal{P}$; i.e., $\max_{x_c, r} \; r$, subject to $B(x_c, r) \subseteq \mathcal{P}$.



Example: Optimal Power Assignment in Wireless Communications

- Consider a wireless comm. system with K transmitters & K receivers.



- Receiver i is intended to receive information only from Transmitter i, & it sees the other transmitters as interferers.

- The signal-to-interference-and-noise ratio (SINR) at receiver i

$$\gamma_i = \frac{G_{ii} p_i}{\sum_{j \neq i} G_{ij} p_j + \sigma_i^2}$$

where

p_i is the transmitter i power,

G_{ij} is the path gain from transmitter j to receiver i,

σ_i^2 is the noise power at receiver i.

- **Problem:** Maximize the weakest SINR subject to power constraints $0 \leq p_i \leq p_{\mathsf{max},i}$, where $p_{\mathsf{max},i}$ is the max. allowable power of transmitter i.

$$\max_{\substack{p_i \in [0, p_{\max,i}] \\ i=1,\ldots,K}} \min_{i=1,\ldots,K} \frac{G_{ii} p_i}{\sum_{j \neq i} G_{ij} p_j + \sigma_i^2}$$

Example: Filter Design in Digital Signal Processing

Finite impulse response (FIR) filter

$$y[k] = \sum_{i=0}^{n-1} h_i x[k-i]$$

where

$x[k]$ input sequence

$y[k]$ output sequence

h_i filter coefficients

n filter order or length

Frequency response:

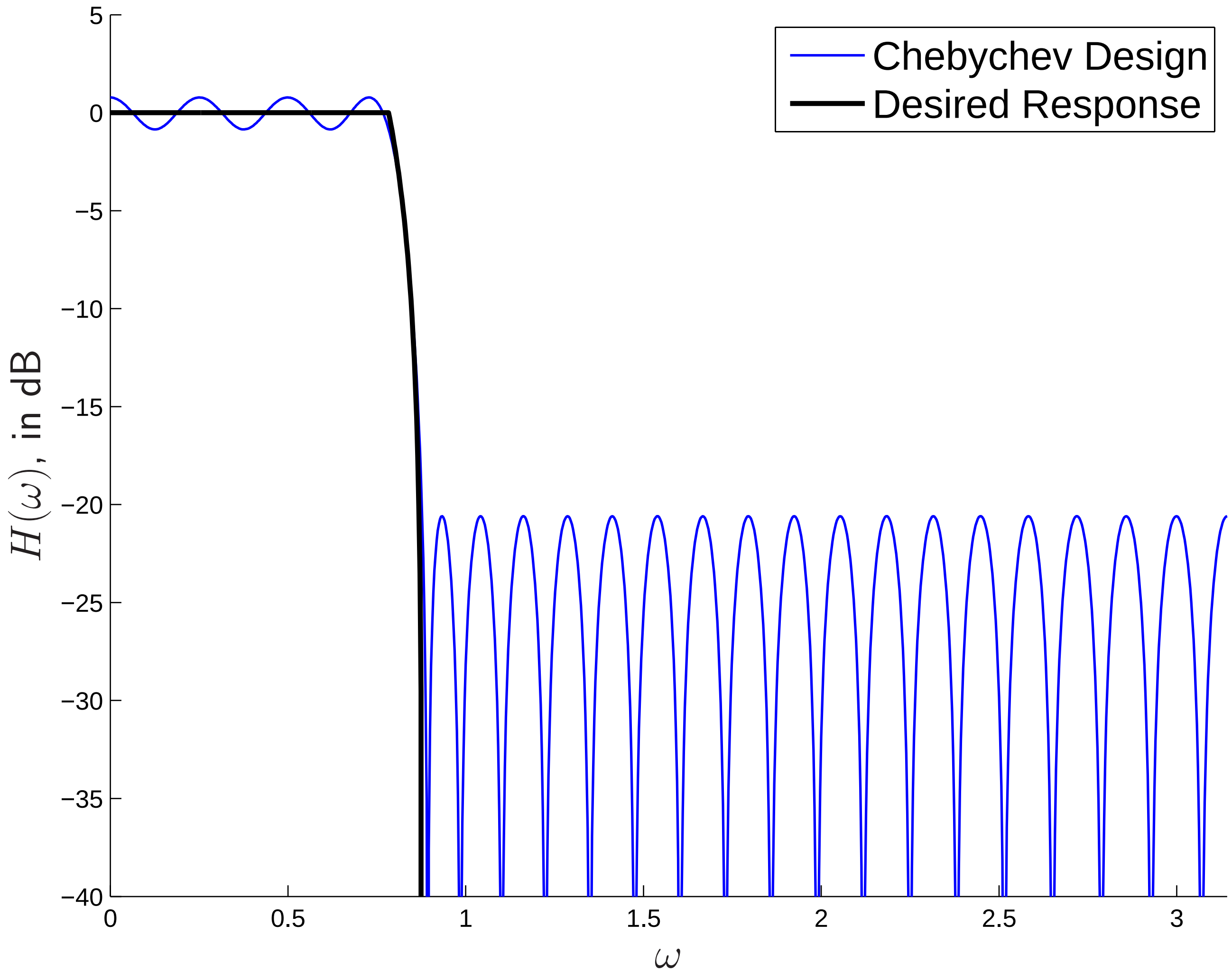
$$H(\omega) = \sum_{i=0}^{n-1} h_i e^{-j\omega i}$$

Problem: find $h = [\ h_0, \ldots, h_{n-1}\ ]^T$ so that h and/or H satisfy/optimize certain given specifications.

For example, the Chebychev design solves for

$$\min_{h \in \mathbf{R}^n} \ \max_{\omega \in [0,\pi]} \left| H(\omega) - H_{\rm des}(\omega) \right|$$

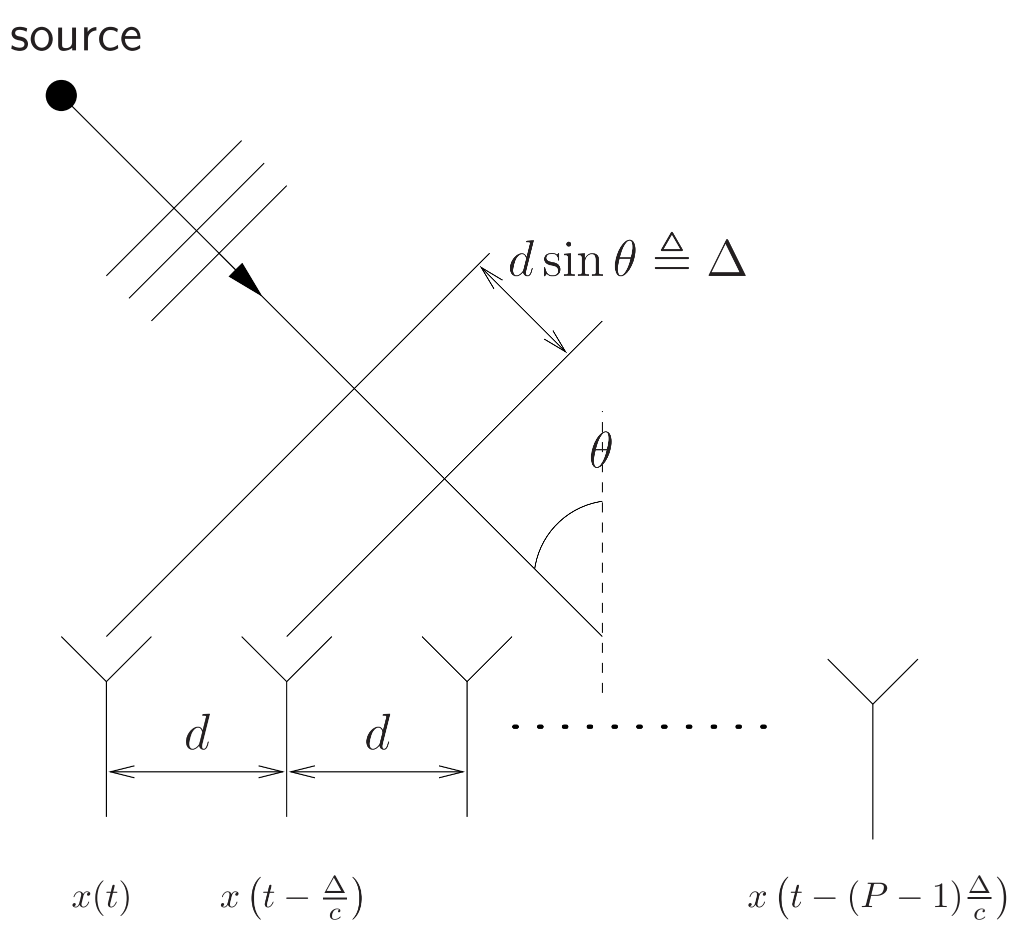
given a desired freq. response $H_{\rm des}(\omega)$. This design minimizes the worst-case absolute error between the desired and actual freq. responses.



With modern software and with proper training in convex optimization, you can programme this filter design in 30 minutes.

Example: Beamformer Design in Array Signal Processing

Uniform linear array:



Signal model:

A1) far-field situations so that source waves are planar; &

A2) narrowband source signals so that the received signal of one sensor is a phase shifted version of that of another.

If a source signal $s(t) \in \mathbf{C}$ comes from a direction of θ, the array output $y(t) = [\ y_1(t), \ldots, y_P(t)\ ]^T$ is

$$y(t) = a(\theta)s(t)$$

Here,

$$a(\theta) = [\ 1, e^{-j2\pi d \sin(\theta)/\lambda}, \ldots, e^{-j2\pi d(P-1)\sin(\theta)/\lambda}\ ]^T \in \mathbf{C}^P$$

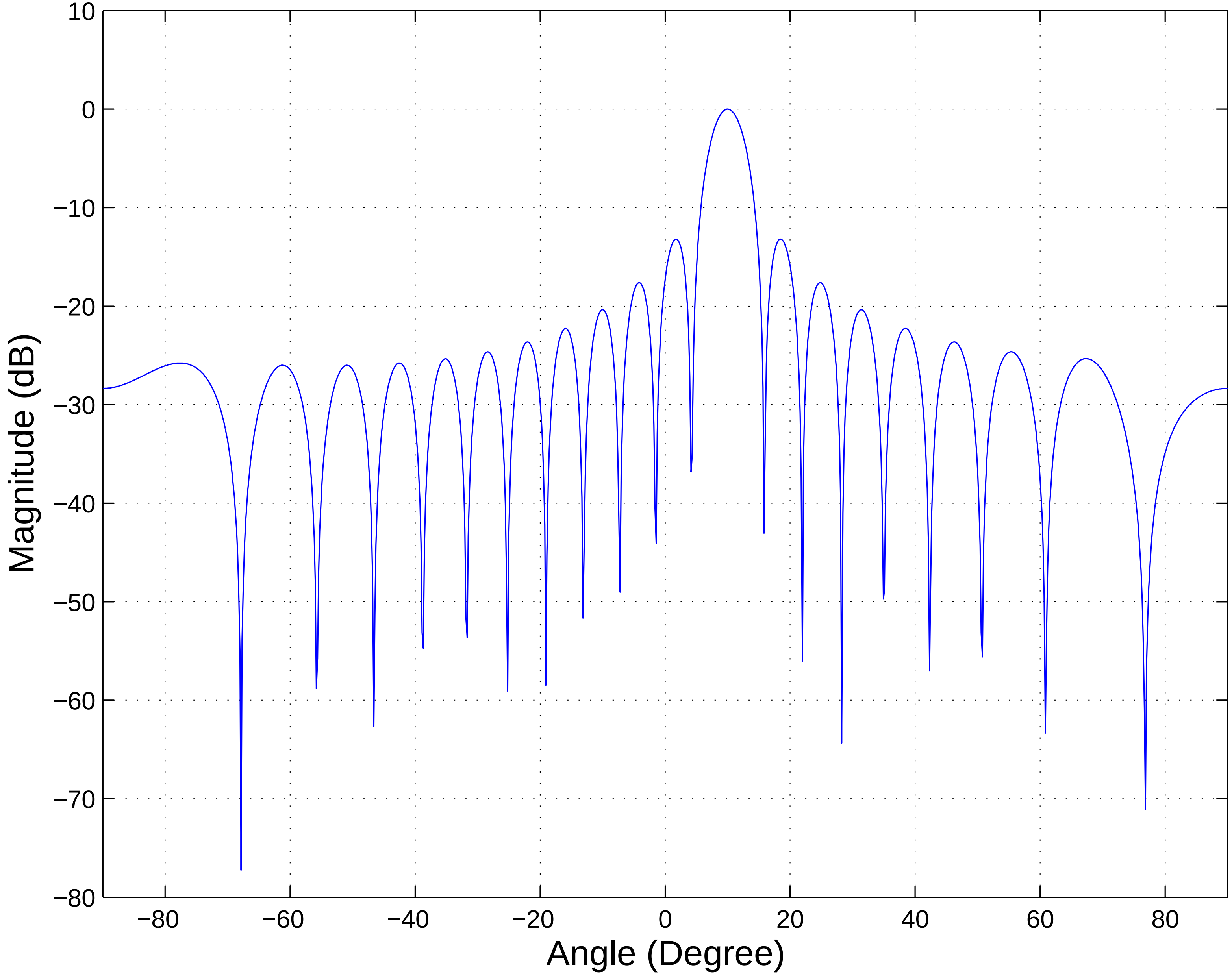
is the steering vector, where λ is the signal wavelength.

Beamforming:

$$\hat{s}(t) = w^H y(t)$$

where $w \in \mathbf{C}^P$ is a beamformer weight vector.

- Let $\theta_{\text{des}} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ be the desired direction.
- A simple beamformer is $w = a(\theta_{\text{des}})$, but it does not provide good sidelobe suppression.
- **Problem:** find a w which minimizes sidelobe energy subject to a pass response to θ_{des}.



Direction pattern of the conventional beamformer. $\theta_{\text{des}} = 10^\circ$; $P = 20$.

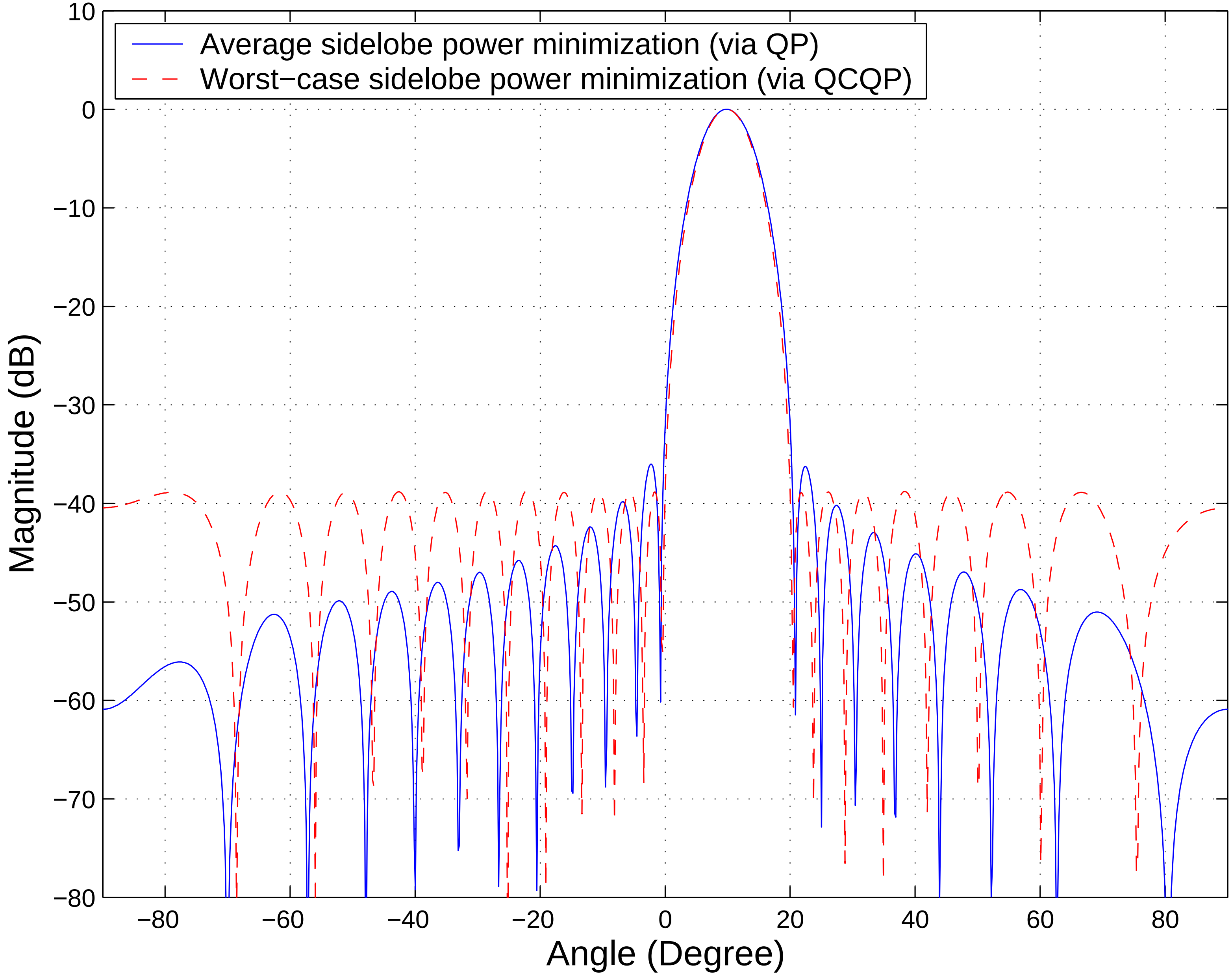
- Let $\Omega = [-\pi/2, \theta_\ell] \cup [\theta_u, \pi/2]$ be the sidelobe band, where $[\theta_\ell, \theta_u]$ is the band of interest.

- Average sidelobe energy minimization:

$$\begin{aligned} \min \ & \int_\Omega |w^H a(\theta)|^2 d\theta \\ \text{s.t.} \ & w^H a(\theta_{\mathsf{des}}) = 1 \end{aligned}$$

- Worst-case sidelobe energy minimization:

$$\begin{aligned} \min \ & \max_{\theta \in \Omega} |w^H a(\theta)|^2 \\ \text{s.t.} \ & w^H a(\theta_{\mathsf{des}}) = 1 \end{aligned}$$



Direction patterns of the two beamformer designs. $\theta_{\text{des}} = 10^\circ$. Sidelobe suppression is applied to directions outside $[0^\circ, 20^\circ]$.

Example: L1 Signal Reconstruction

- Consider a noiseless model

$$y = Ax$$

but with $m < n$ (underdetermined system).

- By linear algebra, there are many (infinite) possible solutions to $y = Ax$

$$Ax = y \iff x = A^\dagger y + u, \; u \in \mathcal{N}(A)$$

where $A^\dagger = A^T(AA^T)^{-1}$, and $\mathcal{N}(A) = \{u \mid Au = 0\}$ is the nullspace of A.

- How can we choose x from these possible solutions?

- This is what matrix textbooks would tell us to do: **Least 2-norm:**

$$\begin{aligned} \min \quad & \|x\|_2 \\ \text{s.t.} \quad & Ax = y \end{aligned}$$

The solution has a closed form, $x^\star = A^\dagger y$ (make sense intuitively).

- **Least 0-norm reconstruction:**

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & Ax = y \end{aligned}$$

  where $\|x\|_0$ counts the number of nonzero elements in x.

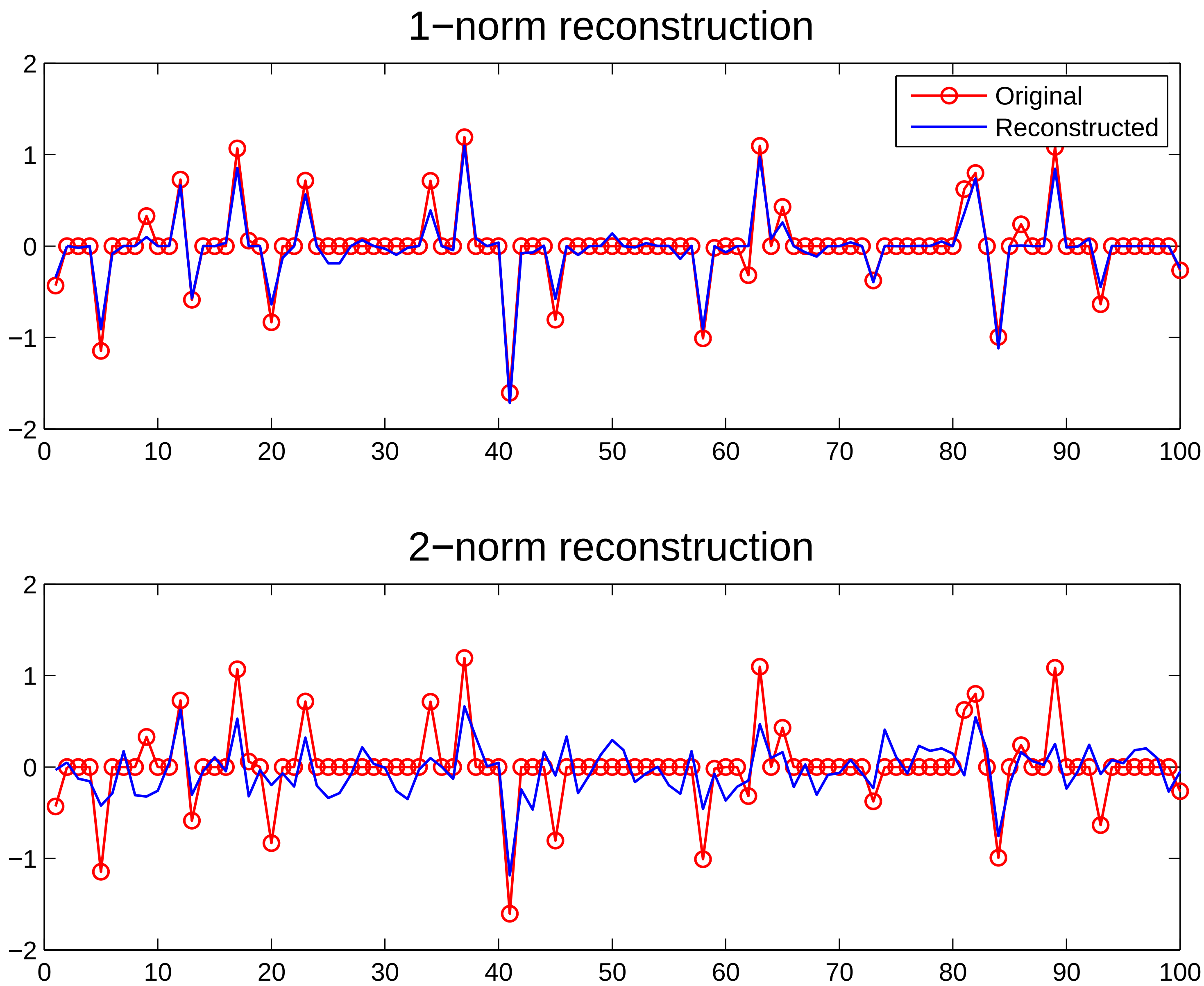
- Make sense for sparse signals; i.e., signals with many zeros.
- Can prove that if the no. of zeros in the actual x is sufficiently large compared to the no. of measurements m, then 0-norm minimization leads to the ground truth.
- $\|\cdot\|_0$ is not convex. In fact, 0-norm minimization poses a very hard problem.

- **Least 1-norm reconstruction:**

$$\begin{array}{rl} \min & \|x\|_1 \\ \text{s.t.} & Ax = y \end{array}$$

- the 'best' convex approximation to 0-norm.
- can prove that under some assumptions, 1-norm minimization is able to approach 0-norm minimization (in some probabilistic sense).
- Currently a very hot topic (in the literature it is called compressive sensing)
- 1-norm minimization is an LP

$$\begin{array}{rl} \min & \sum_{i=1}^n t_i \\ \text{s.t.} & Ax = y \\ & -t_i \leq x_i \leq t_i, \quad i = 1, \ldots, n \end{array}$$


1−norm reconstruction
Original
Reconstructed
2
1
0
−1
−2
0
10
20
30
40
50
60
70
80
90
100
2−norm reconstruction
2
1
0
−1
−2
0
10
20
30
40
50
60
70
80
90
100

Solving optimization problems

- Given a general opt. problem, obtaining its optimal solution can be very hard.
- In the study of *nonlinear programming* (or nonlinear opt.), various numerical algorithms have been developed to *try to* find the optimal solution. Some well known examples are the gradient descent method, and the Newton method.
- These opt. algs. require an initial guess of the solution. Also, they may only guarantee convergence to a locally optimal solution.

- There are also a variety of approaches for nonlinear opt., such as the heuristics based approach (e.g., genetic algs., ant colony opt.), and the Monte-Carlo based approach (e.g., simulated annealing). Again, convergence to a globally optimal solution is not guaranteed.
- When an opt. problem is combinatorial (or discrete), the problem is generally much harder to solve; study complexity theory for the details.
- There are certain problem classes that can be solved effectively, though.

There are problems for which the solutions can be analytically found:

Example: Least Squares (LS)

$$\min \|Ax - b\|_2^2$$

- LS has a closed form solution $x^\star = (A^T A)^{-1} A^T b$.

Example: Entropy Maximization

- Let y be a r.v. drawn from $\{y_1, \ldots, y_n\}$.
- Let $p_i = \mathbf{prob}(y = y_i)$.
- **Problem:** find a distribution $\{p_i\}$ such that the entropy of y is maximized.

$$
\begin{aligned}
\max \ & \textstyle\sum_{i=1}^n p_i \log(1/p_i) \\
\text{s.t.} \ & p_i \geq 0, \quad i = 1, \ldots, n \\
& \textstyle\sum_{i=1}^n p_i = 1
\end{aligned}
$$

- The solution is well known to be $p_i = 1/n$ for all i.

- Opt. problems that can be analytically solved are considered very special cases.
- Another problem class that can be effectively handled is that of the **convex optimization** problems.

Example: Linear Programming (LP)

$$\begin{aligned} &\min\ c^T x \\ &\text{s.t. } a_i^T x \le b_i, \quad i = 1, \ldots, m \end{aligned}$$

- no analytical formula for the solution
- efficient & reliable algorithms for finding the optimal solution exist.

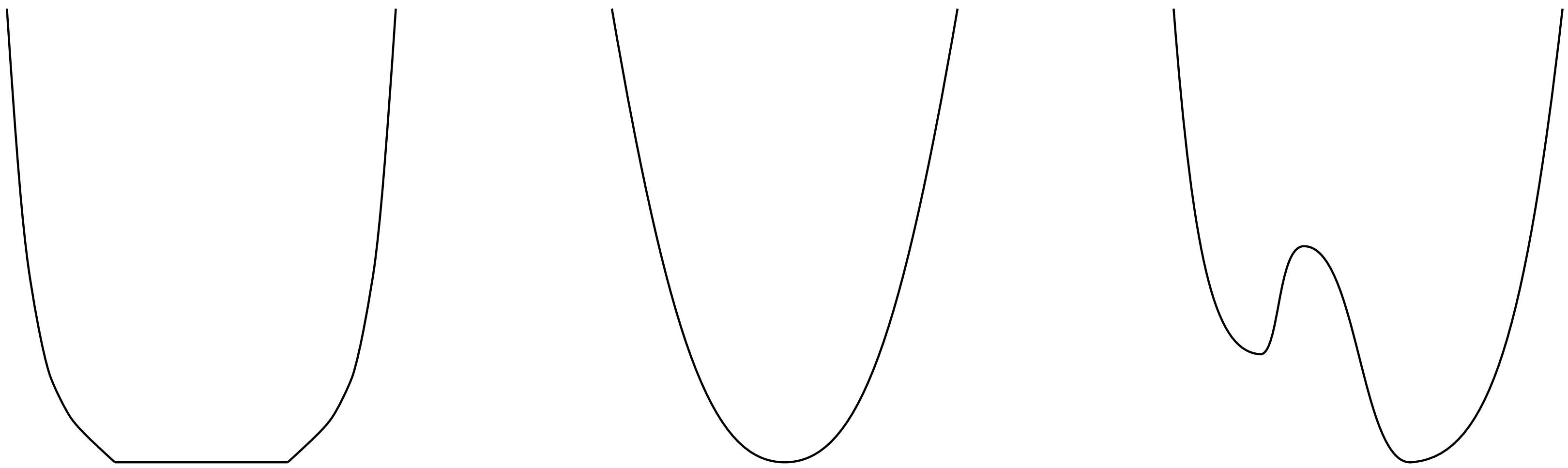
Convex Optimization Problems

$$\begin{aligned} &\min \ f_0(x) \\ &\text{s.t.} \ f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{aligned}$$

in which the objective & constraint functions are convex:

$$f_i(\alpha x + (1-\alpha)y) \leq \alpha f_i(x) + (1-\alpha) f_i(y)$$

for any x, y, and for any $\alpha \in [0, 1]$.



convex strictly convex non-convex

Conic Problems

Conic optimization is a representative class of convex optimization problems.

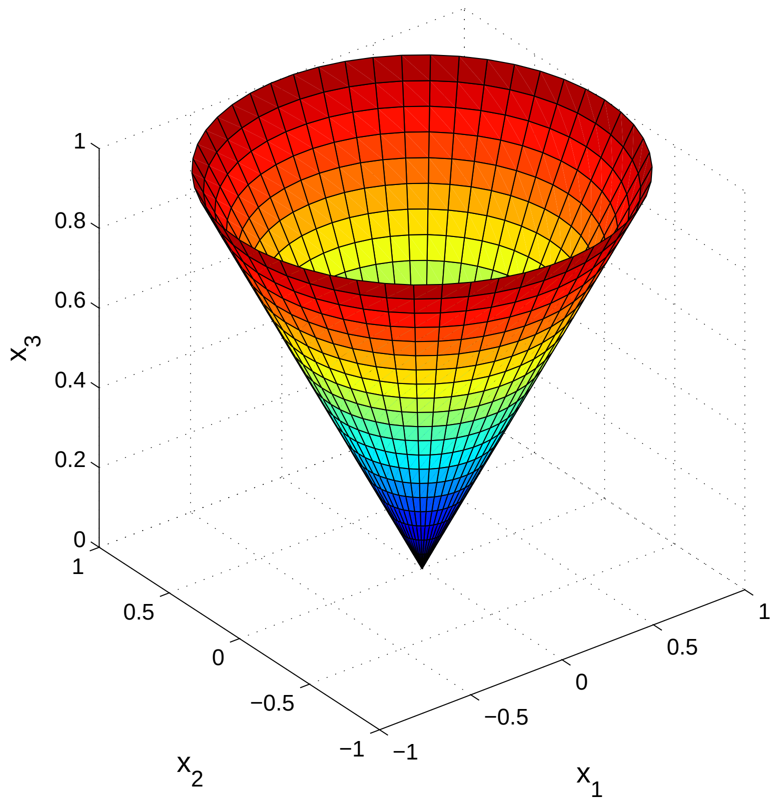
$$\begin{aligned} \min\ & c^T x \\ \text{s.t. } & Ax = b, \\ & x \in \mathcal{K} \end{aligned}$$

where $\mathcal{K}$ is a **convex cone**.

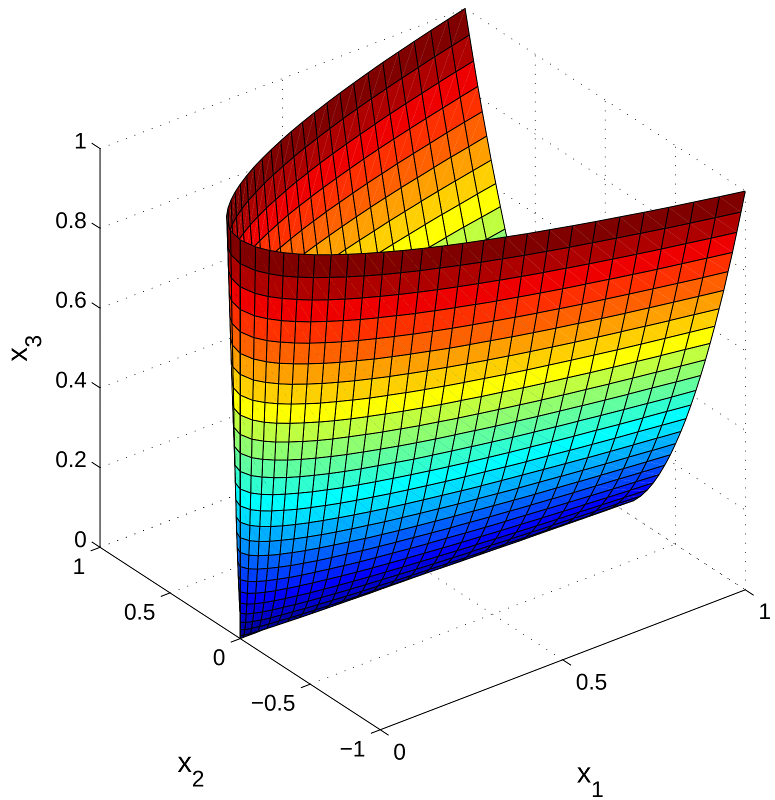
The well-known linear program is conic, where $\mathcal{K} = \{ x \in \mathbf{R}^n \mid x_i \geq 0, i = 1, \ldots, n \}$.

Second order cone program: $\mathcal{K} = \{ \, x \in \mathbf{R}^{n+1} \mid \sqrt{\sum_{i=1}^{n} x_i^2} \leq x_{n+1} \, \}$ is the second order cone.

Semidefinite program: $\mathcal{K} = \{ \, X \in \mathbf{R}^{n \times n} \mid X$ is positive semidefinite (PSD) $\}$ is the set of PSD matrices.



(a) Second-order cone



(b) PSD cone

Merits of convex optimization

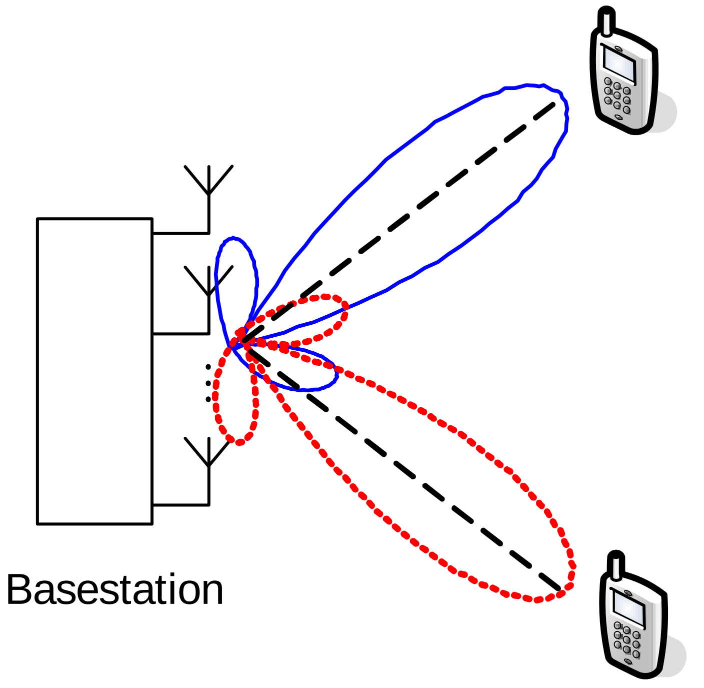
- Reliable and efficient algorithms exist for *many* convex opt. problems, especially those under the conic opt. problem class.
- (Surprisingly) many problems can be converted to convex opt.

Role of convex optimization in nonconvex problems

- We can use convex opt. to approximate a nonconvex problem.
  (Having a hard problem does not mean you should give up).
- We can use convex opt. to build a nonconvex opt. algorithm (subopt. *per se*).
  - SQP, a general-purpose nonlinear opt. solver in MATLAB, may be seen as an algorithm that uses convex opt. to sequentially process an opt. problem.

Example: Unicast Transmit Downlink Beamforming

Problem scenario: a multi-antenna basestation sends data streams, one for each user.



- Transmit signal:

$$x(t) = \sum_{i=1}^{K} w_i s_i(t)$$

where $s_i(t) \in \mathbb{C}$ is the data stream for user i, & $w_i \in \mathbb{C}^N$ its tx beamformer vector.

- Received signal of user i:

$$\begin{aligned} y_i(t) &= h_i^T x(t) + v_i(t) \\ &= h_i^T w_i s_i(t) + \underbrace{\sum_{l \neq i} h_i^T w_l s_l(t)}_{\text{interference}} + v_i(t) \end{aligned}$$

- Let $R_i = \mathrm{E}\{h_i^* h_i^T\}$ be the channel correlation. Assuming that only the channel corr. info. is available, the SINR of user i is

$$\mathsf{SINR}_i = \frac{w_i^H R_i w_i}{\sum_{l \neq i} w_l^H R_i w_l + \sigma_i^2}$$

- The problem is to design $\{w_i\}_{i=1}^k$ such that the total power is minimized whilst all users have SINRs no worse than a requirement γ:

$$\begin{aligned} \min_{w_1,\ldots,w_K \in \mathbb{C}^N} \quad & \textstyle\sum_{i=1}^K \|w_i\|^2 \\ \text{s.t.} \quad & \frac{w_i^H R_i w_i}{\sum_{l \neq i} w_l^H R_i w_l + \sigma_i^2} \geq \gamma, \quad i = 1, \ldots, K \end{aligned}$$

- This problem is nonconvex, but can be solved optimally by a semidefinite program.

Example: Multicast Transmit Beamforming

Problem scenario: a multi-antenna basestation broadcasts common information to all users.

- Received signal model for user i

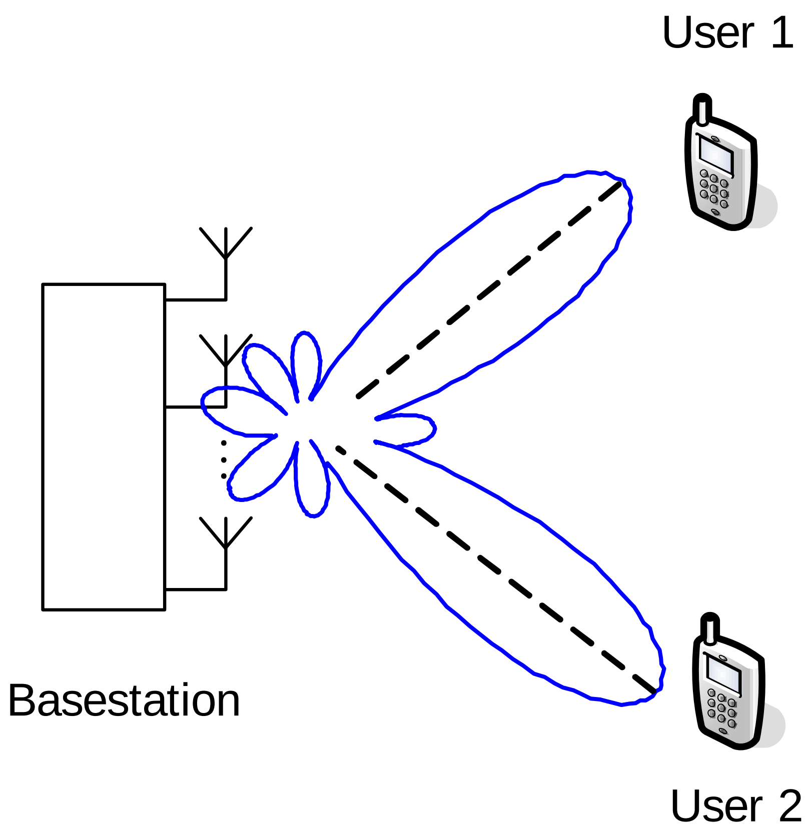
$$y_i(t) = h_i^T x(t) + v_i(t)$$

where h_i is the MISO channel from the basestation to user i, $v_i(t)$ is noise with variance σ_i^2.

- The transmit signal:

$$x(t) = ws(t)$$

where $s(t) \in \mathbb{C}$ is the transmitted common information, & $w \in \mathbb{C}^N$ is the transmit beamformer vector.



- The design problem is to minimize the transmit power subject to SNR constraints:

$$\begin{aligned} \min_{w \in \mathbb{C}^N} \quad & \|w\|^2 \\ \text{s.t.} \quad & \frac{\mathrm{E}\{|h_i^T w s(t)|^2\}}{\sigma_i^2} \geq \gamma, \quad i = 1, \ldots, K \end{aligned}$$

where γ is pre-specified.

- The design problem can be rewritten as a complex-valued nonconvex QCQP

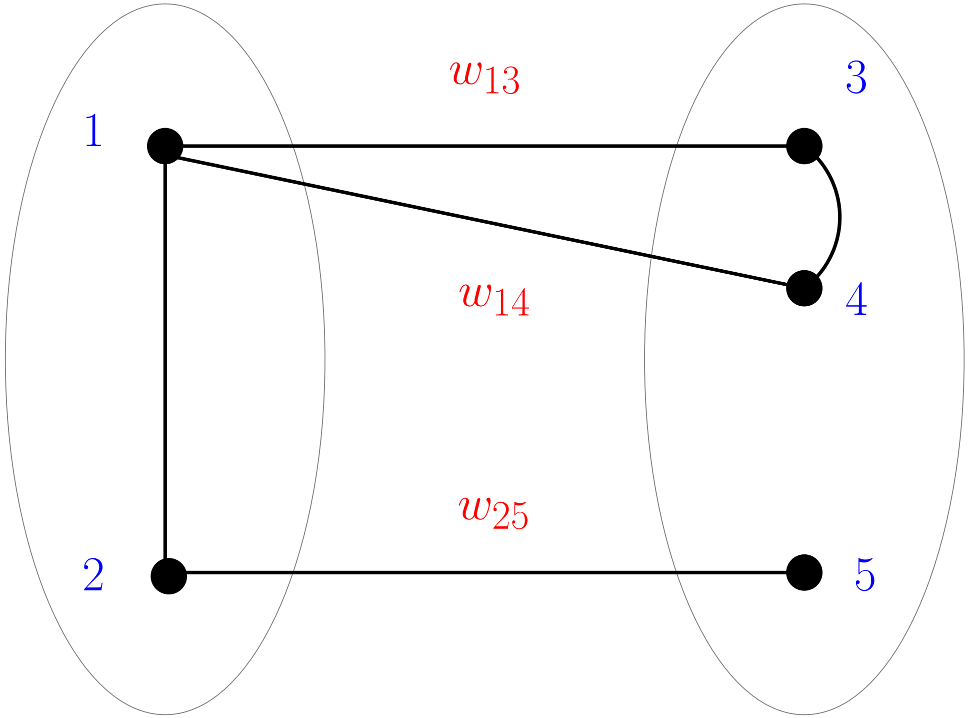
$$\begin{aligned} \min \quad & \|w\|^2 \\ \text{s.t.} \quad & w^H A_i w \geq 1, \quad i = 1, \ldots, K \end{aligned}$$

where $A_i = \mathrm{E}\{h_i^* h_i^T\}/\gamma\sigma_i^2$.

- This multicast transmit beamforming problem is NP-hard in general, but can be approximated by semidefinite programming.

Example: MAXCUT

- **Input:** A graph $G = (V, E)$ with weights w_{ij} for $(i, j) \in E$. Assume $w_{ij} \geq 0$ and $w_{ij} = 0$ if $(i, j) \notin E$.
- **Goal:** Divide nodes into two parts so as to maximize the weight of the edges whose nodes are in different parts.



Let $V = \{1, 2, \ldots, n\}$ without loss of generality. The MAXCUT problem takes the form

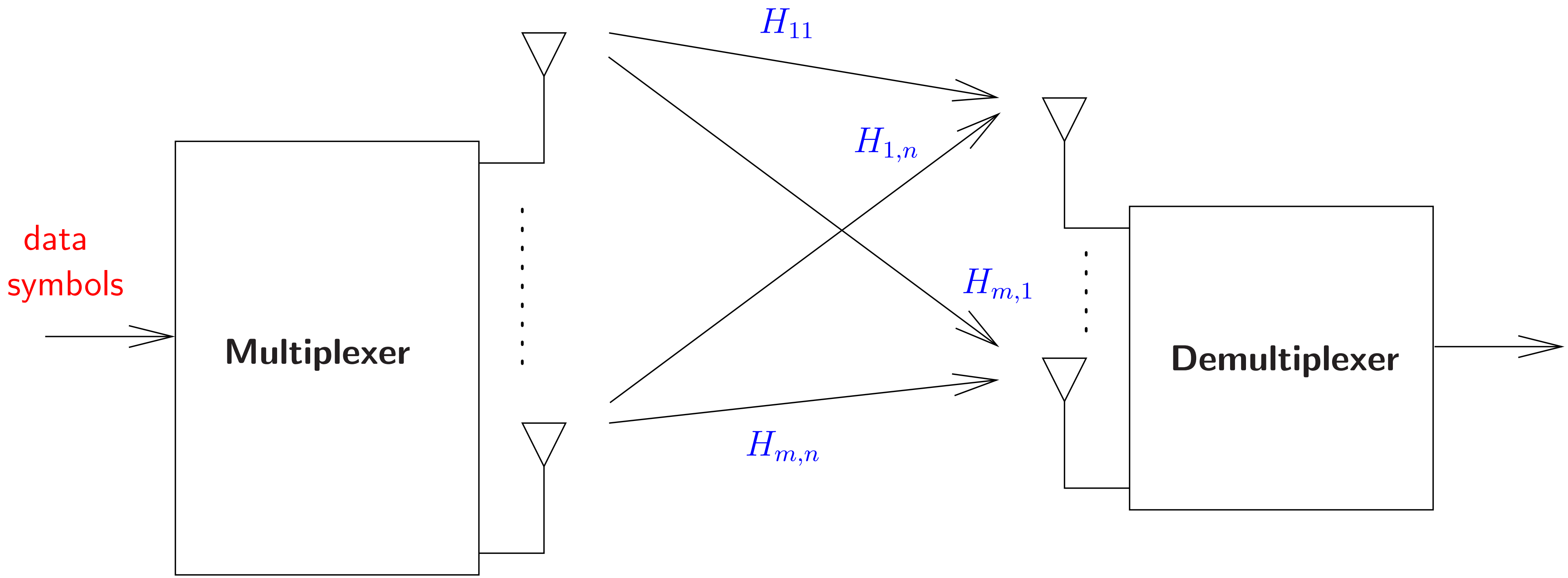
$$\begin{aligned} \max \ & \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \frac{1 - x_i x_j}{2} \\ \text{s.t.} \ & x_i \in \{-1, +1\}, \quad i = 1, \ldots, n \end{aligned}$$

Some remarks:

- MAXCUT is a combinatorial opt. problem.
- To find the optimal solution of MAXCUT is very hard.

Example: MIMO detection

- Consider a multi-antenna communications system with n transmit & m receive antennas.



- The received signal model may be expressed as

$$y = Hs + v$$

where

$y \in \mathbf{R}^m$	multi-receiver output vector;
$s \in \{-1, +1\}^n$	transmitted symbols;
$H \in \mathbf{R}^{m \times n}$	MIMO (or multi-antenna) channel;
$v \in \mathbf{R}^n$	Gaussian noise with zero mean & covariance $\sigma^2 I$.

- Maximum-likelihood (ML) detection of s:

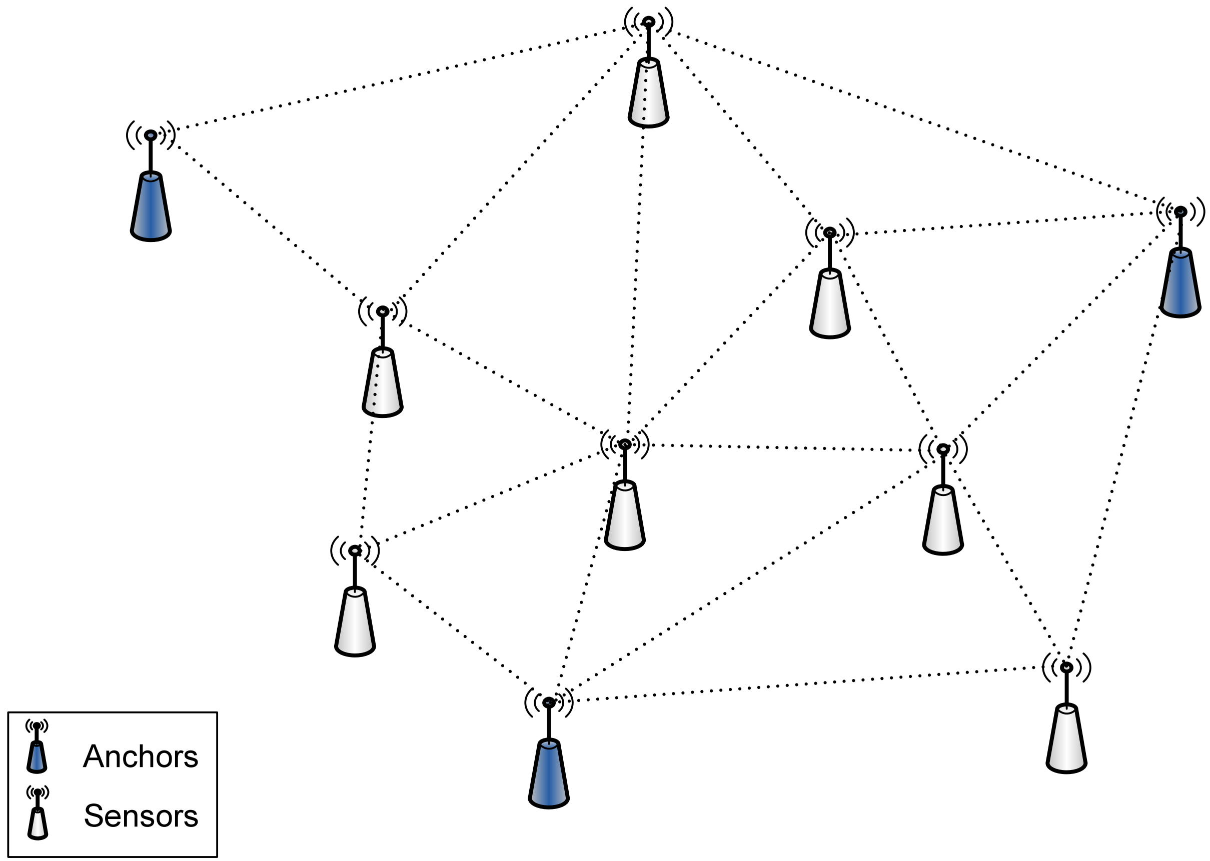
$$\min_{s \in \{\pm 1\}^n} \|y - Hs\|_2^2$$

The ML detection problem is no easier to solve than MAXCUT.

- Nevertheless, ML MIMO detection and MAXCUT can be very effectively approximated by **semidefinite programming**.

Example: Sensor Network Localization

- Consider a sensor network where each node can communicate with the other nodes.
- Depending on applications, a sensor node may have no information on its position.



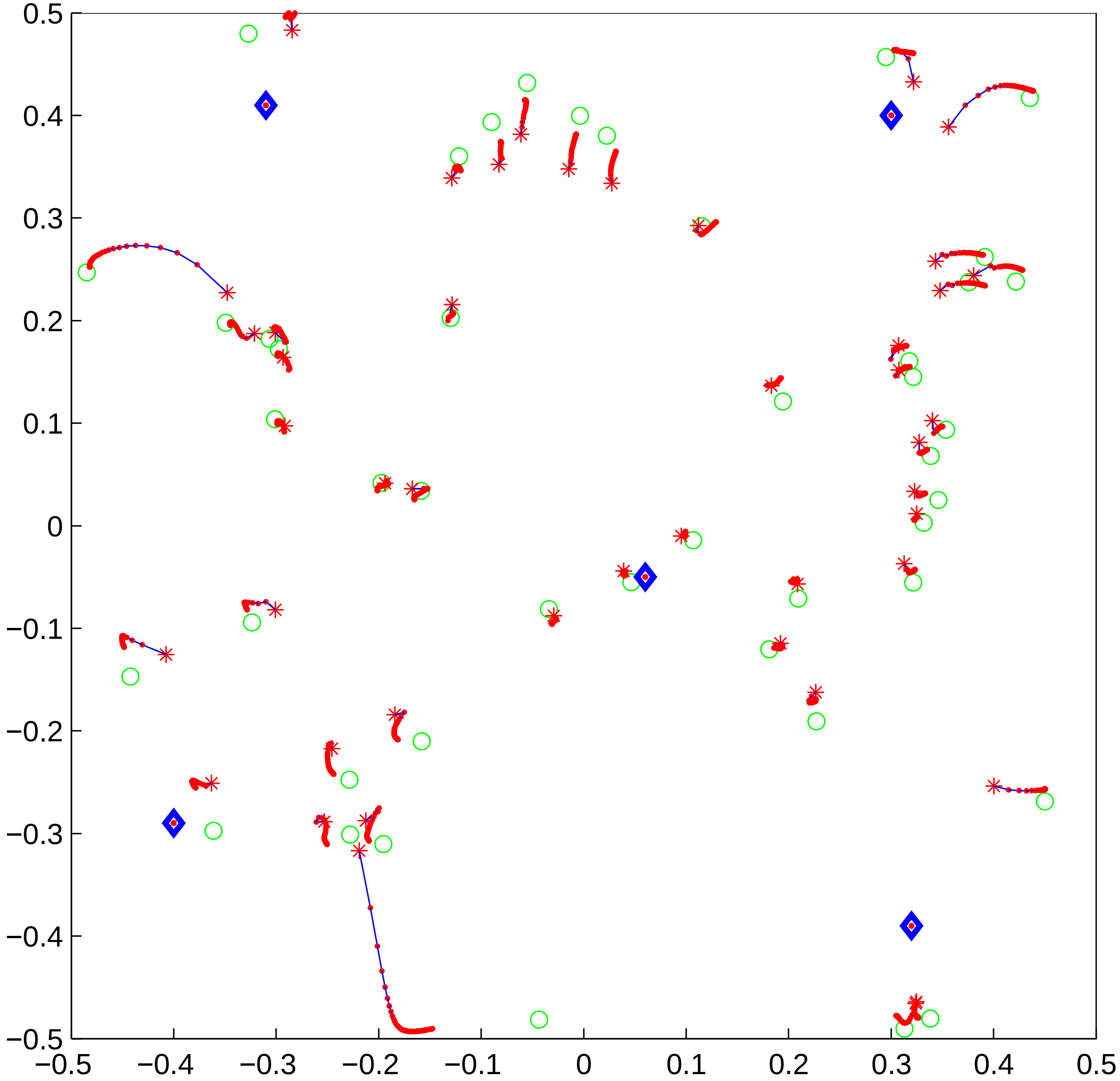
- Let $z_i = [\ x_i, y_i\ ]^T$ be the position of i sensor node.
- A node can communicate with a neighboring node to obtain a distance information

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

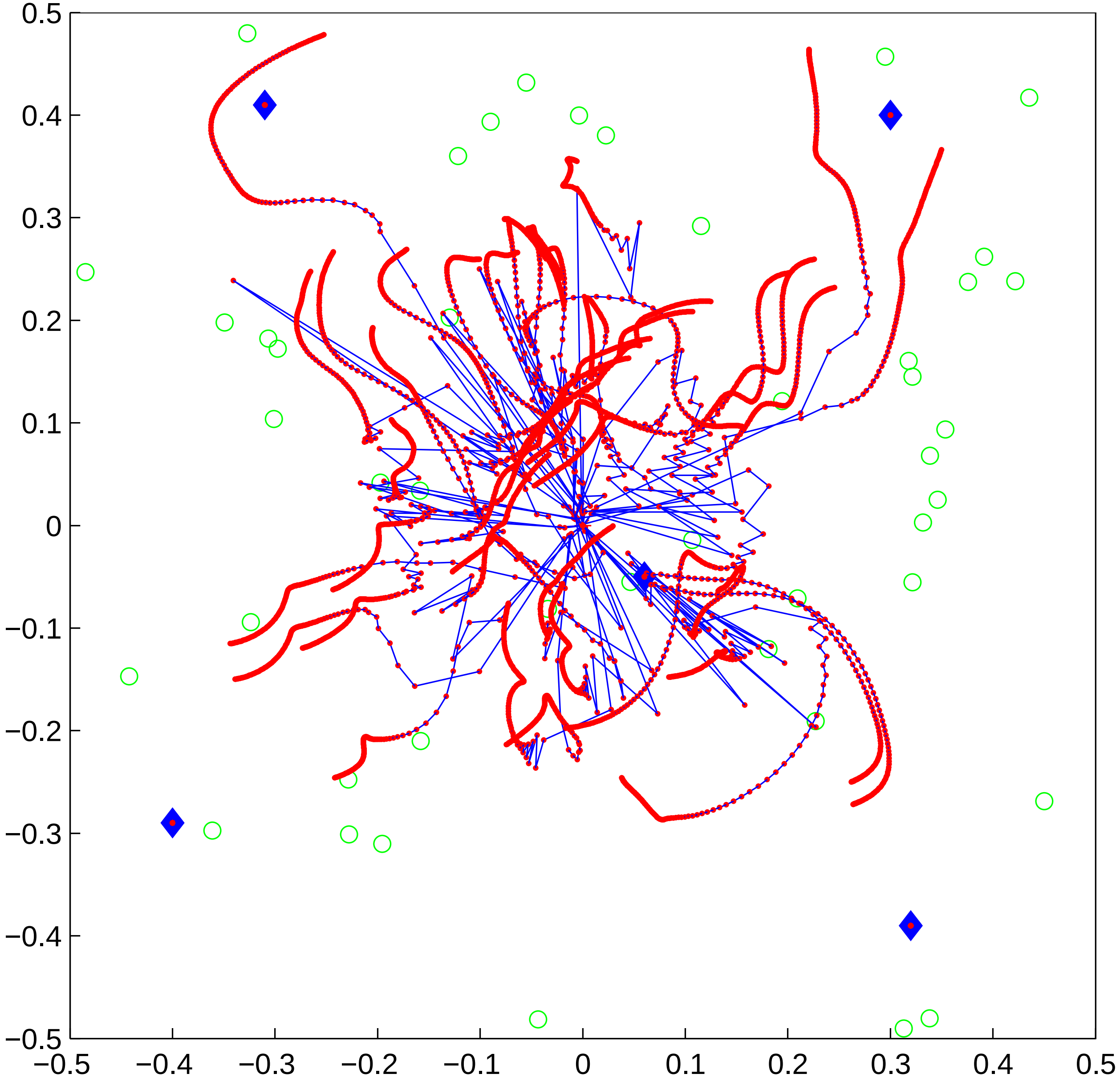
- A sensor network localization problem:

$$\min_{z_i \in \mathbf{R}^2,\ i=1,\ldots,m} \sum_{i,j} \left(d_{ij} - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \right)^2$$

Such a problem is very difficult to solve. But it can be effectively approximated by semidefinite programming.



SDR (ML-SNL formulation), plus a 2nd-stage solution refinement by gradient descent. The distance measurements are noisy. ○: true sensor locations; ◇: anchor locations; ∗: SDR solution; —— : gradient descent trajectory (50 iterations).



Gradient descent ML-SNL with a random starting point. ○: true sensor locations; ◇: anchor locations; —— : gradient descent trajectory (50 iterations).

Course Outline

- Theory:
  - linear algebra and matrix analysis
  - convex sets and convex functions
  - convex optimization, duality
  - conic opt.: linear program, second-order cone program, semidefinite program
- Methods:
  - interior-point methods (a brief overview)
  - subgradient methods, first-order optimization
  - nonconvex problems: optimization strategies based on convex or tractable opt.

- Applications:
  - signal and image processing: optimization of digital filters, optimization of beamforming in sensor array processing, sparse optimization and compressive sensing
  - signal estimation: signal recovery or regression via ℓ_p norm optimization
  - pattern classification: support vector machine, large-margin nearest-neighbor classification
  - wireless communications: power allocation in wireless networks, MIMO transmit optimization, mobile positioning and sensor network localization
  - information theory: capacity optimization in Gaussian frequency-selective and MIMO channels
  - circuit: geometric program and circuit optimization
  - distributed optimization, with applications to image processing and wireless networks

Assessment Method

- Assignments, 20%.
- A written exam., 30%, in midterm. To make sure that you understand the basics.
- Project, 50%. Let you have hands-on experience with using opt. to solve signal processing or engineering problems. Problems will be provided. You can also propose problems related to your own research (subject to my approval).

Course Information

- Venue and Time: Every Wed, 19:00-22:00, ERB405
- Course website: `http://dsp.ee.cuhk.edu.hk/eleg5481/`
- My email: wkma@ee.cuhk.edu.hk

References

Textbook

S. Boyd & L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.
Available online: `http://www.stanford.edu/~boyd/cvxbook/`

Other references

- D. P. Bertsekas, *et. al*, *Convex Analysis and Optimization*, Athenta Scientific, 2002.
- R. Fletcher, *Practical Methods of Opt.*, John Wiley & Sons, 1987.
- `http://www.stanford.edu/class/ee364/`