Optimisation Combinatoire et Convexe.

Approximation results
Today

- Semidefinite relaxations
- Lagrangian relaxations for general QCQPs
- Randomization
- Bounds on suboptimality (MAXCUT)
- Exact relaxations, $S$-lemma
- Concentration arguments
- Approximate $S$-lemma
- Problems on graphs
Convex Optimization

Convex problem:

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_1(x) \leq 0, \ldots, f_m(x) \leq 0 \\
\end{align*}
\]

\(x \in \mathbb{R}^n\) is optimization variable; \(f_i : \mathbb{R}^n \to \mathbb{R}\) are convex:

\[
f_i(\lambda x + (1 - \lambda)y) \leq \lambda f_i(x) + (1 - \lambda)f_i(y)
\]

for all \(x, y, 0 \leq \lambda \leq 1\)

- includes LS, LP, QP, and many others
- like LS, LP, and QP, convex problems are fundamentally tractable
Nonconvex Problems

Nonconvexity makes problems essentially untractable...

- Sometimes the result of bad problem formulation
- Natural limitation: fixed transaction costs, binary communications, ...

What can be done?... Use convex optimization results to

- Get exact solutions in rare cases.
- Find bounds on the optimal value, by relaxation.
- Get ”good” feasible points via randomization.
Nonconvex Problems

- Focus first on a specific class of problems: general QCQPs
- Large range of applications...

A generic QCQP can be written

$$\begin{align*}
\text{minimize} & \quad x^T P_0 x + q_0^T x + r_0 \\
\text{subject to} & \quad x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m
\end{align*}$$

- If all $P_i$ are p.s.d., this is a convex problem...
- We suppose at least one $P_i$ is not p.s.d.
Example: Boolean Least Squares

Boolean least-squares problem:

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|^2 \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

- basic problem in digital communications
- could check all $2^n$ possible values of $x$ . . .
- an NP-hard problem, and very hard in practice
- many heuristics for approximate solution
two-way partitioning problem described in §5.1.4 of the textbook

\[
\begin{align*}
\text{minimize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

where \( W \in \mathbb{S}^n \), with \( W_{ii} = 0 \).

- a feasible \( x \) corresponds to the partition

\[
\{1, \ldots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}
\]

- the matrix coefficient \( W_{ij} \) can be interpreted as the cost of having the elements \( i \) and \( j \) in the same partition.

- the objective is to find the partition with least total cost

- classic particular instance: MAXCUT (\( W_{ij} \geq 0 \))
Convex Relaxation

The original QCQP

\[
\begin{align*}
\text{minimize} & \quad x^T P_0 x + q_0^T x + r_0 \\
\text{subject to} & \quad x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

can be bounded by, after writing \( X = xx^T \),

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(XP_0) + q_0^T x + r_0 \\
\text{subject to} & \quad \text{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\
& \quad X \succeq xx^T \\
& \quad \text{Rank}(X) = 1
\end{align*}
\]

the only nonconvex constraint is now \( \text{Rank}(X) = 1 \)...
Convex Relaxation: Semidefinite Relaxation

- We can directly relax this last constraint, i.e. drop the nonconvex \( \text{Rank}(X) = 1 \) to keep only \( X \succeq xx^T \)

- The resulting program gives a lower bound on the optimal value

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(XP_0) + q_0^T x + r_0 \\
\text{subject to} & \quad \text{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\
& \quad X \succeq xx^T
\end{align*}
\]

Tricky. . . Can be improved?
Lagrangian Relaxation

From the original problem

\[
\begin{align*}
\text{minimize} & \quad x^T P_0 x + q_0^T x + r_0 \\
\text{subject to} & \quad x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

We can form the Lagrangian:

\[
L(x, \lambda) = x^T \left( P_0 + \sum_{i=1}^{m} \lambda_i P_i \right) x + \left( q_0 + \sum_{i=1}^{m} \lambda_i q_i \right)^T x + r_0 + \sum_{i=1}^{m} \lambda_i r_i
\]

in the variables \( x \in \mathbb{R}^n \) and \( \lambda \in \mathbb{R}_+^m \).
Lagrangian Relaxation: Lagrangian

The dual can be computed explicitly as an (unconstrained) quadratic minimization problem, with:

\[
\inf_{x \in \mathbb{R}} x^T P x + q^T x + r = \begin{cases}
  r - \frac{1}{4} q^T P^\dagger q, & \text{if } P \succeq 0 \text{ and } q \in \mathcal{R}(P) \\
  -\infty, & \text{otherwise}
\end{cases}
\]

we have:

\[
\inf_x L(x, \lambda) = -\frac{1}{4} (q_0 + \sum_{i=1}^{m} \lambda_i q_i)^T (P_0 + \sum_{i=1}^{m} \lambda_i P_i)^\dagger (q_0 + \sum_{i=1}^{m} \lambda_i q_i) + \sum_{i=1}^{m} \lambda_i r_i + r_0
\]

where we recognize a Schur complement...
The dual of the QCQP is then given by:

\[
\begin{align*}
\text{maximize} & \quad \gamma + \sum_{i=1}^{m} \lambda_i r_i + r_0 \\
\text{subject to} & \quad \begin{bmatrix} (P_0 + \sum_{i=1}^{m} \lambda_i P_i) & (q_0 + \sum_{i=1}^{m} \lambda_i q_i) / 2 \\
(q_0 + \sum_{i=1}^{m} \lambda_i q_i)^T / 2 & -\gamma \end{bmatrix} \succeq 0 \\
\lambda_i & \geq 0, \quad i = 1, \ldots, m
\end{align*}
\]

which is a semidefinite program in the variable \( \lambda \in \mathbb{R}^m \) and can be solved efficiently.

Use semidefinite duality to compute the dual of this last program?
Taking the dual again, we get an SDP is given by

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(XP_0) + q_0^T x + r_0 \\
\text{subject to} & \quad \text{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\
& \quad \begin{bmatrix} X & x^T \\ x & 1 \end{bmatrix} \succeq 0
\end{align*}
\]

in the variables \( X \in \mathbb{S}^n \) and \( x \in \mathbb{R}^n \)

- This is a convex relaxation of the original program
- We have recovered the semidefinite relaxation in an “automatic” way
Using the previous technique, we can relax the original Boolean LS problem

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|^2 \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

and relax it as an SDP

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(AX) + 2b^TAx + b^Tb \\
\text{subject to} & \quad \begin{bmatrix} X & x^T \\ x & 1 \end{bmatrix} \succeq 0 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n
\end{align*}
\]

this program then produces a lower bound on the optimal value of the original Boolean LS program
Lagrangian Relaxation: Partitioning

The partitioning problem defined above is

\[
\begin{align*}
\text{minimize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

the variable \( x \) disappears from the relaxation, which becomes

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(W X) \\
\text{subject to} & \quad X \succeq 0 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n
\end{align*}
\]
Feasible points?

- Lagrangian relaxations only provide lower bounds on the optimal value
- Can we compute good feasible points?
- Can we measure how suboptimal this lower bound is?
The original QCQP

\[
\begin{align*}
\text{minimize} & \quad x^T P_0 x + q_0^T x + r_0 \\
\text{subject to} & \quad x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

was relaxed into

\[
\begin{align*}
\text{minimize} & \quad \text{Tr}(X P_0) + q_0^T x + r_0 \\
\text{subject to} & \quad \text{Tr}(X P_i) + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\
& \quad \begin{bmatrix} X & x^T \\
x & 1 \end{bmatrix} \succeq 0
\end{align*}
\]

- The last (Schur complement) constraint is equivalent to $X - xx^T \succeq 0$
- Hence, if $x$ and $X$ are the solution to the relaxed program, then $X - xx^T$ is a covariance matrix...
Randomization

- Pick $x$ as a Gaussian variable with $x \sim \mathcal{N}(x, X - xx^T)$
- $x$ will solve the QCQP "on average" over this distribution

In other words, it will satisfy

\[
\begin{align*}
\text{minimize} & \quad \mathbb{E}[x^T P_0 x + q_0^T x + r_0] \\
\text{subject to} & \quad \mathbb{E}[x^T P_i x + q_i^T x + r_i] \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

A good feasible point can then be obtained by sampling enough $x$. . .
Consider the constraint
\[ x^T Px + q^T x + r \leq 0 \]
we decompose the matrix \( P \) into its positive and negative parts
\[ P = P_+ - P_-, \quad P_+, \; P_- \succeq 0 \]
and original constraint becomes
\[ x^T P_+ x + q_0^T x + r_0 \leq x^T P_- x \]
Linearization

Both sides of the inequality are now convex quadratic functions. We linearize the right hand side around an initial feasible point $x_0$ to obtain

$$x^T P_+ x + q_0^T x + r_0 \leq x^{(0)T} P_- x^{(0)} + 2 x^{(0)T} P_- (x - x^{(0)})$$

- The right hand side is now an affine lower bound on the original function $x^T P_- x$ (see §3.1.3 in the book).
- The resulting constraint is convex and more conservative than the original one, hence the feasible set of the new problem will be a convex subset of the original feasible set.
- We form a *convex restriction* of the problem.

We can then solve the convex restriction to get a better feasible point $x^{(1)}$ and iterate. . .
Bounds on suboptimality

- In certain particular cases, it is possible to get a hard bound on the gap between the optimal value and the relaxation result.
- A classical example is that of the MAXCUT bound.

The MAXCUT problem is a particular case of the partitioning problem:

\[
\begin{align*}
\text{maximize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

its Lagrangian relaxation is computed as:

\[
\begin{align*}
\text{maximize} & \quad \text{Tr}(W X) \\
\text{subject to} & \quad X \succeq 0 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n
\end{align*}
\]
Bounds on suboptimality: MAXCUT

Let $X$ be a solution to this program

- we look for a feasible point by sampling a normal distribution $\mathcal{N}(0, X)$
- we convert each sample point $x$ to a feasible point by rounding it to the nearest value in $\{-1, 1\}$, i.e. taking
  \[
  \hat{x} = \text{sgn}(x)
  \]

crucially, when $\hat{x}$ is sampled using that procedure, the expected value of the objective $\mathbb{E}[\hat{x}^T W x]$ can be computed explicitly:

\[
\mathbb{E}[\hat{x}^T W x] = \frac{2}{\pi} \sum_{i,j=1}^{n} W_{ij} \arcsin(X_{ij}) = \frac{2}{\pi} \text{Tr}(W \arcsin(X))
\]
■ We are guaranteed to reach this expected value $\frac{2}{\pi} \text{Tr}(W \arcsin(X))$ after sampling a few (feasible) points $\hat{x}$

■ Hence we know that the optimal value $OPT$ of the MAXCUT problem is between $\frac{2}{\pi} \text{Tr}(W \arcsin(X))$ and $\text{Tr}(WX)$

Furthermore, with $\arcsin(X) \succeq X$, we can simplify (and relax) the above expression to get:

$$\frac{2}{\pi} \text{Tr}(WX) \leq OPT \leq \text{Tr}(WX)$$

the procedure detailed above guarantees that we can find a feasible point at most $\frac{2}{\pi}$ suboptimal
Bounds on suboptimality: MAXCUT

**Proposition**

**MAXCUT approximation.** Let $OPT$ be the optimal value of the partitioning problem

$$\begin{align*}
\text{maximize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}$$

where $W \succeq 0$, and let $SDP$ be the optimal value of its Lagrangian relaxation

$$\begin{align*}
\text{maximize} & \quad \text{Tr}(W X) \\
\text{subject to} & \quad X \succeq 0 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n
\end{align*}$$

then we have

$$\frac{2}{\pi} \text{Tr}(WX) \leq OPT \leq \text{Tr}(WX).$$

**Proof.** Suppose we sample $x \sim \mathcal{N}(0, X)$ then take $\hat{x} = \text{sgn}(x)$. We get

$$\mathbb{E}[\hat{x}^T W x] = \frac{2}{\pi} \sum_{i,j=1}^{n} W_{ij} \arcsin(X_{ij}) = \frac{2}{\pi} \text{Tr}(W \arcsin(X)).$$
where \( \text{arcsin}(X) \) is taken elementwise, with

\[
\text{arcsin}(X)_{ij} = X_{ij} + \sum_{k=1}^{\infty} \frac{1 \cdot 3 \cdots (2k - 1)}{2^k k!(2k + 1)} X_{ij}^{2k+1}
\]

which means

\[
\text{arcsin}(X) - X \succeq 0
\]

because if we define the **elementwise matrix power** \([X]^k\) such that

\[
[X]^k_{ij} = X_{ij}
\]

then \([X]^k \succeq 0\) when \(X \succeq 0\). This finally means that

\[
E[\hat{x}^T W x] = \frac{2}{\pi} \text{Tr}(W \text{arcsin}(X)) \geq \frac{2}{\pi} \text{Tr}(W X) \quad \blacksquare
\]
Numerical Example: Boolean LS

Boolean least-squares problem:

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|^2 \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

with

\[
\|Ax - b\|^2 = x^T A^T A x - 2 b^T A x + b^T b = \text{Tr} A^T A X - 2 b^T A^T x + b^T b
\]

where \( X = xx^T \), hence can express BLS as

\[
\begin{align*}
\text{minimize} & \quad \text{Tr} A^T A X - 2 b^T A x + b^T b \\
\text{subject to} & \quad X_{ii} = 1, \quad X \succeq xx^T, \quad \text{rank}(X) = 1
\end{align*}
\]

\ldots still a very hard problem
Using Lagrangian relaxation, with

\[ X \succeq xx^T \iff \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \]

we obtained the **SDP relaxation** (with variables \( X, x \))

\[
\begin{align*}
\text{minimize} & \quad \text{Tr} A^T A X - 2b^T A^T x + b^T b \\
\text{subject to} & \quad X_{ii} = 1, \quad \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0
\end{align*}
\]

- Optimal value of SDP gives **lower bound** for BLS
- If optimal matrix is rank one, we’re done
Interpretation via randomization

- Can think of variables $X, x$ in SDP relaxation as defining a normal distribution $z \sim \mathcal{N}(x, X - xx^T)$, with $\mathbb{E} z_i^2 = 1$
- SDP objective is $\mathbb{E} \|Az - b\|^2$

Suggests randomized method for BLS:

- Find $X^{opt}, x^{opt}$, optimal for SDP relaxation
- Generate $z$ from $\mathcal{N}(x^{opt}, X^{opt} - x^{opt}x^{opt^T})$
- Take $x = \text{sgn}(z)$ as approximate solution of BLS (can repeat many times and take best one)
Example

- (randomly chosen) parameters $A \in \mathbb{R}^{150 \times 100}$, $b \in \mathbb{R}^{150}$
- $x \in \mathbb{R}^{100}$, so feasible set has $2^{100} \approx 10^{30}$ points

**LS approximate solution:** minimize $\|Ax - b\|$ s.t. $\|x\|^2 = n$, then round yields objective 8.7% over SDP relaxation bound

**randomized method:** (using SDP optimal distribution)

- best of 20 samples: 3.1% over SDP bound
- best of 1000 samples: 2.6% over SDP bound
Example: Partitioning Problem

\[ \frac{\|Ax - b\|}{(SDP \text{ bound})} \]

<table>
<thead>
<tr>
<th>frequency</th>
<th>SDP bound</th>
<th>LS solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.20</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

A. d’Aspremont. M1 ENS.
Example: Partitioning Problem

**MAXCUT.** Numerical example.

\[
\begin{align*}
\text{minimize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

the Lagrange dual of this problem is given by the SDP:

\[
\begin{align*}
\text{maximize} & \quad -1^T \nu \\
\text{subject to} & \quad W + \text{diag}(\nu) \succeq 0
\end{align*}
\]
Partitioning: Lagrangian relaxation

the dual of this SDP is another SDP

\[
\begin{align*}
\text{minimize} & \quad \text{Tr } W X \\
\text{subject to} & \quad X \succeq 0 \\
& \quad X_{ii} = 1, \quad i = 1, \ldots, n
\end{align*}
\]

the solution \( X^{opt} \) gives a lower bound on the optimal value \( p^{opt} \) of the partitioning problem

- solve the previous SDP to find \( X^{opt} \) (and the bound \( p^{opt} \))
- let \( v \) denote an eigenvector of \( X^{opt} \) associated with its largest eigenvalue
- now let

\[ \hat{x} = \text{sgn}(v) \]

the vector \( \hat{x} \) is our guess for a good partition
Randomization.

- We generate independent samples $x^{(1)}, \ldots, x^{(K)}$ from a normal distribution with zero mean and covariance $X^{\text{opt}}$.
- For each sample we consider the heuristic approximate solution
  $$\hat{x}^{(k)} = \text{sgn}(x^{(k)})$$
- We then take the one with lowest cost.

On a randomly chosen problem:

- The optimal SDP lower bound $p^{\text{opt}}$ is equal to $-1641$.
- The simple $\text{sgn}(x)$ heuristic gives a partition with total cost $-1280$.

At this point, we can say that the optimal value is between $-1641$ and $-1280$. 
Partitioning: Numerical Example

Histogram of the objective obtained by the randomized heuristic, over 1000 samples: the minimum value reached here is $-1328$
We know the optimum is between $-1641$ and $-1328$. 
Greedy method

We can improve these results a little bit using the following simple greedy heuristic.

- Suppose the matrix \( Y = \hat{x}^T W \hat{x} \) has a column \( j \) whose sum \( \sum_{i=1}^{n} y_{ij} \) is positive.
- Switching \( \hat{x}_j \) to \( -\hat{x}_j \) will decrease the objective by \( 2 \sum_{i=1}^{n} y_{ij} \).
- if we pick the column \( y_{j_0} \) with largest sum, switch \( \hat{x}_{j_0} \) and repeat until all column sums are negative, we decrease the objective.

Applying this to the SDP heuristic gives an objective value of \(-1372\), our best partition yet...
Hidden convexity, $S$-lemma, . . .
In general, nonconvex quadratically constrained quadratic programming is hard.

Yet, we have very efficient, very reliable algorithms to solve the following eigenvalue problem
\[
\begin{align*}
\text{maximize} & \quad x^T Ax \\
\text{subject to} & \quad x^T x = 1
\end{align*}
\]
with complexity $O(n^2)$ (computing a sequence of matrix vector products).

Why is this one easy?

**S-lemma.** SDP relaxations of Nonconvex QPs with one quadratic constraint (two in some cases) are exact, hence these programs can be solved in polynomial time.
Geometrically, the set \((x^T Ax, x^T Bx)\), where \(x \in \mathbb{R}^n\), describes a **convex cone**. This has important consequences for semidefinite relaxations.

### Proposition

**Quadratic convexity.** Suppose \(A, B \in S_n\), then for all \(X \succeq 0\), there exists \(x \in \mathbb{R}^n\) with

\[
x^T Ax = \text{Tr}(AX) \quad \text{and} \quad x^T Bx = \text{Tr}(BX).
\] (1)

**Proof.** Suppose it is true for all \(X \in S^n_+\) with \(2 \leq \text{Rank} X \leq k\), i.e. there exists an \(x\) such that (1) holds. Let us show that it also holds if \(\text{Rank} X = k + 1\).

A matrix \(X \in S^n_+\) with \(\text{Rank} X = k + 1\) can be expressed as \(X = yy^T + Z\) where \(y \neq 0\) and \(Z \in S^n_+\) with \(\text{Rank} Z = k\). By assumption, there exists a \(z\) such that \(\text{Tr}(AZ) = z^T Az, \text{Tr}(AZ) = z^T Bz\). Therefore

\[
\text{Tr}(AX) = \text{Tr}(A(yy^T + zz^T)), \quad \text{Tr}(BX) = \text{Tr}(B(yy^T + zz^T)).
\]

\(yy^T + zz^T\) has rank one or two, hence (1) by assumption.
It is therefore sufficient to prove the result if \textbf{Rank} $X = 2$. If \textbf{Rank} $X = 2$, we can factor $X$ as $X = VV^T$ where $V \in \mathbb{R}^{n \times 2}$, with linearly independent columns $v_1$ and $v_2$.

Without loss of generality we can assume that $V^T AV$ is diagonal. (If $V^T AV$ is not diagonal we replace $V$ with $VP$ where $V^T AV = P \text{diag}(\lambda) P^T$ is the eigenvalue decomposition of $V^T AV$.) We will write $V^T AV$ and $V^T BV$ as

$$
V^T AV = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad V^T BV = \begin{bmatrix} \sigma_1 & \gamma \\ \gamma & \sigma_2 \end{bmatrix},
$$

and define

$$
w = \begin{bmatrix} \text{Tr}(AX) \\ \text{Tr}(BX) \end{bmatrix} = \begin{bmatrix} \lambda_1 + \lambda_2 \\ \sigma_1 + \sigma_2 \end{bmatrix}.
$$

We need to show that $w = (x^T Ax, x^T Bx)$ for some $x$. We distinguish two cases.
First, assume \((0, \gamma)\) is a linear combination of the vectors \((\lambda_1, \sigma_1)\) and \((\lambda_2, \sigma_2)\):

\[
0 = z_1 \lambda_1 + z_2 \lambda_2, \quad \gamma = z_1 \sigma_1 + z_2 \sigma_2,
\]

for some \(z_1, z_2\). In this case we choose \(x = \alpha v_1 + \beta v_2\), where \(\alpha\) and \(\beta\) are determined by solving two quadratic equations in two variables

\[
\alpha^2 + 2\alpha \beta z_1 = 1, \quad \beta^2 + 2\alpha \beta z_2 = 1. \tag{2}
\]

This will give the desired result, since

\[
\begin{bmatrix}
(\alpha v_1 + \beta v_2)^T A (\alpha v_1 + \beta v_2) \\
(\alpha v_1 + \beta v_2)^T B (\alpha v_1 + \beta v_2)
\end{bmatrix}
= \alpha^2 \begin{bmatrix}
\lambda_1 \\
\sigma_1
\end{bmatrix}
+ 2\alpha \beta \begin{bmatrix}
0 \\
\gamma
\end{bmatrix}
+ \beta^2 \begin{bmatrix}
\lambda_2 \\
\sigma_2
\end{bmatrix}
= (\alpha^2 + 2\alpha \beta z_1) \begin{bmatrix}
\lambda_1 \\
\sigma_1
\end{bmatrix}
+ (\beta^2 + 2\alpha \beta z_2) \begin{bmatrix}
\lambda_2 \\
\sigma_2
\end{bmatrix}
= \begin{bmatrix}
\lambda_1 + \lambda_2 \\
\sigma_1 + \sigma_2
\end{bmatrix}.
\]

It remains to show that the equations (2) are solvable. To see this, we first note
that \( \alpha \) and \( \beta \) must be nonzero, so we can write the equations equivalently as

\[
\alpha^2(1 + 2(\beta/\alpha)z_1) = 1, \quad (\beta/\alpha)^2 + 2(\beta/\alpha)(z_2 - z_1) = 1.
\]

The equation \( t^2 + 2t(z_2 - z_1) = 1 \) has a positive and a negative root. At least one of these roots (the root with the same sign as \( z_1 \)) satisfies \( 1 + 2tz_1 > 0 \), so we can choose

\[
\alpha = \pm 1/\sqrt{1 + 2tz_1}, \quad \beta = t\alpha.
\]

This yields two solutions \((\alpha, \beta)\) that satisfy (2). (If both roots of \( t^2 + 2t(z_2 - z_1) = 1 \) satisfy \( 1 + 2tz_1 > 0 \), we obtain four solutions.)

Next, assume that \((0, \gamma)\) is not a linear combination of \((\lambda_1, \sigma_1)\) and \((\lambda_2, \sigma_2)\). In particular, this means that \((\lambda_1, \sigma_1)\) and \((\lambda_2, \sigma_2)\) are linearly dependent. Therefore their sum \( w = (\lambda_1 + \lambda_2, \sigma_1 + \sigma_2) \) is a nonnegative multiple of \((\lambda_1, \sigma_1)\), or \((\lambda_2, \sigma_2)\), or both. If \( w = \alpha^2(\lambda_1, \sigma_1) \) for some \( \alpha \), we can choose \( x = \alpha v_1 \). If \( w = \beta^2(\lambda_2, \sigma_2) \) for some \( \beta \), we can choose \( x = \beta v_2 \).
S-lemma

Strong duality.

- This shows directly that strong duality holds for

\[
\begin{align*}
\text{maximize} & \quad x^T A x \\
\text{subject to} & \quad x^T B x \leq 0
\end{align*}
\]

since the optimum value of this program is equal to that of its bidual, which is a semidefinite program in \( X \).

- Some extensions of this result are possible, e.g. the inhomogeneous case

\[
\begin{align*}
\text{maximize} & \quad x^T A x + a^T x \\
\text{subject to} & \quad x^T B x + b^T x + c = 0
\end{align*}
\]

or the normalized case (known as Brickman’s theorem)

\[
\begin{align*}
\text{maximize} & \quad x^T A x \\
\text{subject to} & \quad x^T B x \leq 0 \\
& \quad x^T x = 1
\end{align*}
\]
Concentration inequalities, approximate $S$-lemma, etc. . .
Concentration inequalities

- We can extend randomization arguments to constrained problems.
- Concentration inequalities allow us to bound the probability that a constraint is feasible. Basically, if we match the constraints/objective on average, we can find w.h.p. a feasible point whose objective value is not too far off.

Theorem

**Gaussian concentration.** Suppose $f(x) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with constant $L$ with respect to the Euclidean norm, i.e.

\[
|f(y) - f(x)| \leq L \|x - y\|_2, \quad \text{for all } x, y \in \mathbb{R}^n
\]

then if $g_i, i = 1, \ldots, n$ are i.i.d. Gaussian variables with $g_i \sim \mathcal{N}(0, 1)$, we have

\[
\text{Prob} \left[ |M - f(g)| \geq Lt \right] \leq \exp(-t^2/2)
\]

where $M = \mathbb{E}[f(g)]$ or its median.
# Concentration inequalities

Similar concentration results also exist for binary random variables.

## Theorem

**Bernstein inequality.** Let $u_i \in \{-1, 1\}$ be i.i.d. random variables with $\mathbb{E}[u_i] = 0$, for any $a \in \mathbb{R}^n$ we have

$$
\Pr \left[ |a^T u| \geq t \|a\|_2 \right] \leq \exp(-t^2/4)
$$
Approximate $S$-lemma

We can show the following result extending the $S$-lemma to approximate the case with multiple quadratic constraints. (Inhomogeneous extensions are possible).

**Theorem**

**Approximate $S$-lemma.** Call $OPT$ the optimal value of the following quadratic optimization problems

\[
\begin{align*}
\text{maximize} & \quad x^T A x \\
\text{subject to} & \quad x^T A_i x \leq c_i, \quad i = 1, \ldots, m
\end{align*}
\]

in the variable $x \in \mathbb{R}^n$, where the matrix $A \in \mathcal{S}_n$ is arbitrary, $c_i > 0$, and $A_i \succeq 0$. Call $SDP$ the optimal value of the semidefinite program (we assume strong duality holds and $SDP < \infty$)

\[
\begin{align*}
\text{maximize} & \quad \text{Tr}(AX) \\
\text{subject to} & \quad \text{Tr}(A_i X), \quad i = 1, \ldots, m
\end{align*}
\]

in the variable $X \in \mathcal{S}_n$. Then $OPT \leq SDP \leq 2 \ln \left( 2 \sum_{i=1}^m \text{Rank}(A_i) \right) OPT$. 

A. d’Aspremont. M1 ENS.
**Proof.** We write $X$ an optimal solution to SDP and $X^{1/2}AX^{1/2} = UDU^T$, the eigenvalue decomposition of $X^{1/2}AX^{1/2}$, with $D$ diagonal and $U$ orthogonal. We have, by construction

$$\text{Tr}(D) = \text{Tr}(UDU^T) = \text{Tr}(X^{1/2}AX^{1/2}) = \text{Tr}(AX) = SDP$$

We let $\xi_i \in \{-1, 1\}$ be i.i.d. random variables with $E[\xi] = 0$. We define $\eta = X^{1/2}U\xi$, and write $D_i = U^T X^{1/2} A_i X^{1/2} U$, such that

$$\text{Tr}(D_i) = \text{Tr}(U^T X^{1/2} A_i X^{1/2} U) = \text{Tr}(X^{1/2} A_i X^{1/2}) = \text{Tr}(A_i X) \leq c_i$$

this means

$$\eta^T A\eta = \xi^T U^T X^{1/2} A X^{1/2} U \xi$$
$$= \xi^T U^T U D U^T U \xi$$
$$= \xi^T D \xi = \text{Tr}(D) = SDP,$$

and similarly,

$$E[\eta^T A_i \eta] = E[\xi^T U^T X^{1/2} A_i X^{1/2} U \xi]$$
$$= \text{Tr}(U^T X^{1/2} A_i X^{1/2} U) = \text{Tr}(A_i X) \leq c_i$$
This shows that the vector $\eta$ solves the SDP “on average”. We now show how to construct vectors that satisfy approximately solve the QP with high probability. We can write

$$D_i = \sum_{j=1}^{k} d_j d_j^T, \quad k = \text{Rank}(D_i) = \text{Rank}(A_i),$$

Using the previous concentration inequality

$$\text{Prob}[|d_j^T \xi| \geq \sqrt{t}\|d_j\|_2] \leq 2 \exp(-t/2),$$

now, for each given $\xi$, if $\xi^T D_i \xi \geq t \sum_{j=1}^{k} \|d_j\|_2^2$ then for at least for one $j$, we have $|d_j^T \xi| \geq \sqrt{t}\|d_j\|_2$, hence

$$\text{Prob}[\xi^T D_i \xi \geq t \sum_{j=1}^{k} \|d_j\|_2^2] \leq \sum_{i=1}^{k} \text{Prob}[|d_j^T \xi| \geq \sqrt{t}\|d_j\|_2] \leq 2 \text{Rank}(D_i) \exp(-t/2)$$

Now, we have $\sum_{j=1}^{k} \|d_j\|_2^2 = \text{Tr}(\sum_{j=1}^{k} d_j d_j^T) = \text{Tr}(D_i) \leq c_i$. Hence we have showed

$$\text{Prob}[\eta^T A_i \eta \geq tc_i] \leq 2 \text{Rank}(A_i) \exp(-t/2)$$
Let $\delta > 0$ and

$$\Theta = 2 \ln \left( \frac{\sum_{i=1}^{m} \text{Rank}(A_i)}{1 - \delta} \right),$$

using union bounds, with probability $\delta > 0$

$$\Theta^{-1/2} \eta$$

will be a feasible point of the QP, reaching an objective value of $\Theta^{-1} \text{SDP}$, hence

$$\text{OPT} \leq \text{SDP} \leq 2 \ln \left( 2 \sum_{i=1}^{m} \text{Rank}(A_i) \right) \text{OPT} \quad \blacksquare$$