

APPROXIMATE MAXIMUM-LIKELIHOOD ESTIMATION USING SEMIDEFINITE PROGRAMMING

*Joachim Dahl**, *Bernard H. Fleury*

Dept. of Communication Technology
Aalborg University, Frederik Bajersvej 7A
9220 Aalborg, Denmark

Lieven Vandenberghe

Dept. of Electrical Engineering
University of California, Los Angeles
Los Angeles, California 90095-1594, USA

ABSTRACT

We consider semidefinite relaxations of a quadratic optimization problem with polynomial constraints. This is an extension of quadratic problems with boolean variables. Such combinatorial problems can in general not be solved in polynomial time. Semidefinite relaxations has been proposed as a promising technique to give provable good bounds on certain boolean quadratic problems in polynomial time. We formulate the extensions from boolean variables to quaternary variables using (i) a polynomial relaxation or (ii) by using standard semidefinite relaxations of a linear transformation of boolean variables. We analytically compare the two different approaches of relaxation. The relaxations can all be expressed as semidefinite programs, which can be solved efficiently using e.g. interior point methods. Applications of our results include maximum likelihood estimation in communication systems, which we explore in simulations in order to compare the quality of the different relaxations with optimal solutions.

1. INTRODUCTION

In this paper we consider extensions of quadratic problems with boolean variables. The quadratic problem with boolean constraints is well-studied in the literature and has important applications in e.g. graph assignment theory. A boolean quadratic problem has the form

$$\begin{aligned} &\text{minimize} && x^T A x - 2b^T x \\ &\text{subject to} && x_i^2 = 1, \quad i = 1, \dots, n \end{aligned} \quad (1)$$

where $A = A^T \in \mathbf{R}^{n \times n}$ and $b \in \mathbf{R}^n$. This is a combinatorial problem which in general cannot be solved in polynomial time. In this work we investigate semidefinite relaxations of the combinatorial problem (1) and extensions thereof. The semidefinite relaxations yield a bound on the original problem which can be solved in polynomial time. In special cases the bound is even provable good.

Although, the combinatorial problems we consider cannot in general be solved in polynomial time, for certain combinatorial quadratic problems the problem complexity grows polynomially with the problem dimensions. This is e.g. the case for minimum-capacity-cuts for directed graphs or for maximum likelihood estimation in multiuser detection with non-positive crosscorrelations between waveforms. Both problems can be expressed as quadratic problems with boolean constraints, see [1] for further references.

*The author is funded by Analog Devices Incp.

The maximum likelihood application will be explored in more detail in Sec. 4 in our simulations.

We start, however, by reviewing standard results for the boolean quadratic problem. The boolean problem (1) is difficult to solve due to the non-convex constraints $x_i^2 = 1$. Instead we find a lower bound to the problem (1) using the dual function. The dual problem is

$$\begin{aligned} &\text{maximize} && -\mathbf{1}^T \lambda - b^T (A + \mathbf{diag}(\lambda))^\dagger b \\ &\text{subject to} && A + \mathbf{diag}(\lambda) \succeq 0 \\ &&& b \in \mathcal{R}(A + \mathbf{diag}(\lambda)) \end{aligned} \quad (2)$$

in the variable $\lambda \in \mathbf{R}^n$. The dual problem is (always) concave and can in this case be solved as a semidefinite program.

Let $f(x) = x^T A x - 2b^T x$ denote the primal objective of the boolean quadratic problem, and let $g(\lambda)$ denote the dual objective in (2). From weak duality we have that $g(\lambda) \leq f(x)$. However, it is well-known that strong duality fails for this problem, so in general the dual problem (2) gives a lower bound on the original problem (1).

An alternative representation of the dual problem can be obtained by rewriting (2) as

$$\begin{aligned} &\text{maximize} && -\mathbf{1}^T z \\ &\text{subject to} && \begin{bmatrix} A & -b \\ -b^T & 0 \end{bmatrix} + \mathbf{diag}(z) \succeq 0 \end{aligned} \quad (3)$$

with the variable $z \in \mathbf{R}^{n+1}$. The dual problem of (3) is

$$\begin{aligned} &\text{minimize} && \text{Tr} X \begin{bmatrix} A & -b \\ -b^T & 0 \end{bmatrix} \\ &\text{subject to} && X_{ii} = 1, \quad i = 1, \dots, n+1 \\ &&& X \succeq 0. \end{aligned} \quad (4)$$

Slater's constraint qualification states that strong duality of the convex problem (3) holds provided there exists a strictly feasible interior point, which is obviously satisfied. Since strong duality holds, we have that $g(\lambda^*) = h(X^*)$ where $h(X)$ denotes the objective of (4), i.e. problem (2) and problem (4) obtain the same value at their optimal points. We next recast the original boolean quadratic problem (1) as

$$\begin{aligned} &\text{minimize} && \text{Tr} X \begin{bmatrix} A & -b \\ -b^T & 0 \end{bmatrix} \\ &\text{subject to} && X = \begin{bmatrix} x \\ c \end{bmatrix} \begin{bmatrix} x^T & c \end{bmatrix} \\ &&& X_{ii} = 1, \quad i = 1, \dots, n+1. \end{aligned} \quad (5)$$

with the additional variable $c \in \mathbf{R}$. Note that for $c = 1$ the two formulations (1) and (5) are identical. Then (4) follows from (5) if we replace the rank-1 constraint by the positive semidefinite relaxation $X \succeq 0$. This gives an interesting interpretation of the Lagrangian dual problem as a semidefinite rank-relaxation of the original problem.

2. RANDOMIZED ROUNDING

If the solution to (4) has rank 1, the optimal solution is readily observed from X^* . Otherwise we need to map the solution back to a boolean vector. In their seminal paper [2] Goemans and Williamson proposed to perform this mapping using a randomized algorithm. They generate a vector $y \sim \mathcal{N}(0, X^*)$ and estimate x_i according to the rule

$$x_i = \text{sgn}(y_i), \quad i = 1, \dots, n.$$

Under the assumption that the elements of the coefficient matrix (in our case $\begin{bmatrix} A & -b \\ -b^T & 0 \end{bmatrix}$) are all nonnegative, Goemans and Williams *proved* that the expected value of the relaxation using the randomized rounding technique is within 0.878 of the optimal value.

These results were later extended by Nesterov [3] to the general case without assumptions on the coefficient matrix. For this case Nesterov proved that relative accuracy of the semidefinite relaxation is within 0.429 of the optimal solution. Compared to the 0.878 bound by Goemans and Williams the general 0.429 bound seems discouraging, but in practice this worst-case bound is rather conservative, see simulation results in Sec. 4.

To tighten the relaxation the rounding is repeated several times and the realization of x that achieves the largest objective of (1) is chosen as the boolean estimate. The cost of performing the randomized rounding is in most cases insignificant compared to solving the semidefinite relaxations.

3. HIGHER ORDER POLYNOMIAL CONSTRAINTS

We next consider extensions to higher order polynomial constraints of the special form

$$\begin{aligned} & \text{minimize} && x^* A x - 2 \text{Re } b^* x \\ & \text{subject to} && x_i^d = 1, \quad i = 1, \dots, n \end{aligned} \quad (6)$$

where $A = A^* \in \mathbf{C}^{n \times n}$, $b \in \mathbf{C}^n$ and d is a power of two. In solving (6) we are faced with the same difficulty as for the boolean problem (1), namely that the problem is non-convex and in general cannot be solved in polynomial time.

We consider two different semidefinite relaxations of the problem (6). The two methods can thus be interpreted as a straightforward extension of the problem (1) to incorporate complex variables and special polynomial constraints of the form $x_i^d = 1$. The methods described in the sequel extend to arbitrary values of d , but becomes notationally cumbersome for higher values. For purpose of exposition and due to space limitations we limit our discussion to $d = 4$.

3.1. Polynomial relaxation

The constraint $x_i^4 = 1$ implies that $x_i \in \{-1, 1, -j, j\}$. We define the auxiliary variable $y_i = x_i^2$ so that $y_i^2 = 1$. This in

turn implies that $y_i \in \{-1, +1\}$ and $x_i \in \{-1, 1, -j, +j\}$. Let $u = \text{Re } x$ and $v = \text{Im } x$. Then the above conditions are equivalent to

$$\begin{aligned} y_i^2 &= 1 \\ y_i &= u_i^2 - v_i^2 \\ u_i v_i &= 0. \end{aligned}$$

We next recast problem (6) in terms of real variables. For that purpose let $B = \text{Re } A$, $C = \text{Im } A$, $c = \text{Re } b$ and $d = \text{Im } b$. Then an equivalent formulation using real variables is

$$\begin{aligned} & \text{minimize} && \begin{bmatrix} u^T & v^T \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} - 2 \begin{bmatrix} c^T & d^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && u_i^2 - v_i^2 = y_i, \quad i = 1, \dots, n \\ & && u_i v_i = 0, \quad i = 1, \dots, n \\ & && y_i^2 = 1. \end{aligned}$$

We obtain a semidefinite relaxation by reformulating the problem as

$$\begin{aligned} & \text{minimize} && \text{Tr} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} - 2 \begin{bmatrix} c^T & d^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && \text{diag}(X_{11}) - \text{diag}(X_{22}) = y \\ & && \text{diag}(X_{12}) = 0 \\ & && \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \begin{bmatrix} u^T & v^T \end{bmatrix} \\ & && \text{diag}(Y) = 1 \\ & && Y = yy^T \end{aligned}$$

and then relaxing the rank-1 constraints

$$\begin{aligned} & \text{minimize} && \text{Tr} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} - 2 \begin{bmatrix} c^T & d^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && \text{diag}(X_{11}) - \text{diag}(X_{22}) = y \\ & && \text{diag}(X_{12}) = 0 \\ & && \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succeq \begin{bmatrix} u \\ v \end{bmatrix} \begin{bmatrix} u^T & v^T \end{bmatrix} \\ & && \text{diag}(Y) = 1 \\ & && Y \succeq yy^T. \end{aligned}$$

This in turn gives a reduced expression for the relaxation as

$$\begin{aligned} & \text{minimize} && \text{Tr} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} - 2 \begin{bmatrix} c^T & d^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && -1 \preceq \text{diag}(X_{11}) - \text{diag}(X_{22}) \preceq 1 \\ & && \text{diag}(X_{12}) = 0 \\ & && \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succeq \begin{bmatrix} u \\ v \end{bmatrix} \begin{bmatrix} u^T & v^T \end{bmatrix}. \end{aligned} \quad (7)$$

The dual problem of (7) can be written as

$$\begin{aligned} & \text{minimize} && \|y\|_1 + \gamma \\ & \text{subject to} && \begin{bmatrix} B & -C & c \\ C & B & d \\ c^T & d^T & 0 \end{bmatrix} + \begin{bmatrix} \text{diag}(y) & \text{diag}(w) & 0 \\ \text{diag}(w) & \text{diag}(y) & 0 \\ 0 & 0 & \gamma \end{bmatrix} \succeq 0. \end{aligned} \quad (8)$$

3.2. Affine transformation

For many polynomial constraints we can express the variables as an affine transformation of simpler boolean variables. By introducing such a transformation the problem then reduces to a boolean

quadratic problem where we can use the standard semidefinite relaxation of Sec. 1.

We first recast the problem in terms of real variables as

$$\begin{aligned} & \text{minimize} && [u^T \ v^T] \begin{bmatrix} B & -C \\ C & B \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} - 2 [c^T \ d^T] \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && (u_i + v_i)^2 = 1, \quad i = 1, \dots, n \\ & && (u_i - v_i)^2 = 1, \quad i = 1, \dots, n. \end{aligned}$$

We obtain a semidefinite relaxation by first reformulating the problem as

$$\begin{aligned} & \text{minimize} && \text{Tr} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} - 2 [c^T \ d^T] \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && \text{diag} \left(\begin{bmatrix} I & I \\ I & -I \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \right) = 1 \\ & && \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} [u^T \ v^T] \end{aligned}$$

and replacing the rank-1 constraint by a semidefinite constraint

$$\begin{aligned} & \text{minimize} && \text{Tr} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} - 2 [c^T \ d^T] \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && \text{diag} (X_{11} + 2X_{12} + X_{22}) = 1 \\ & && \text{diag} (X_{11} - 2X_{12} + X_{22}) = 1 \\ & && \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} [u^T \ v^T]. \end{aligned}$$

The constraints obviously imply that $\text{diag} (X_{12}) = 0$ so we get a reduced semidefinite relaxation

$$\begin{aligned} & \text{minimize} && \text{Tr} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \begin{bmatrix} B & -C \\ C & B \end{bmatrix} - 2 [c^T \ d^T] \begin{bmatrix} u \\ v \end{bmatrix} \\ & \text{subject to} && \text{diag} (X_{11}) + \text{diag} (X_{22}) = 1 \\ & && \text{diag} (X_{12}) = 0 \\ & && \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succeq \begin{bmatrix} u \\ v \end{bmatrix} [u^T \ v^T]. \end{aligned} \quad (9)$$

which has a dual problem

$$\begin{aligned} & \text{minimize} && \mathbf{1}^T z + \gamma \\ & \text{subject to} && \begin{bmatrix} B & -C & c \\ C & B & d \\ c^T & d^T & 0 \end{bmatrix} + \begin{bmatrix} \text{diag} (z) & \text{diag} (w) & 0 \\ \text{diag} (w) & \text{diag} (z) & 0 \\ 0 & 0 & \gamma \end{bmatrix} \succeq 0. \end{aligned} \quad (10)$$

At this point it is instructive to compare the relaxation in (7) with the relaxation in (9). We infer that the relaxation in (9) is better (tighter) since the constraint

$$\text{diag} (X_{11} + X_{22}) = 1$$

is embedded in the larger constraint set

$$-1 \preceq \text{diag} (X_{11} - X_{22}) \preceq 1.$$

This important and interesting relationship is consistently supported by our simulations in Sec. 4.2. Comparing the two initial relaxations, it is not obvious which is better.

4. MAXIMUM-LIKELIHOOD ESTIMATION

We explore the application of the semidefinite relaxations for maximum likelihood estimation for communication systems. The purpose of this investigation is two-fold; it extends previous results

reported in [4, 5] and it gives a useful scenario for testing the quality of the different relaxations. Polynomial constraints as considered in Sec. 3 have nice interpretations as different memory-less modulation schemes, e.g. the constraint $x_i^M = 1$ corresponds to M -PSK.

In general maximum-likelihood estimation is a very difficult problem. Here, we consider a single user communication systems with a time-dispersive propagation channel modeled as a time-invariant transversal filter. For this specific scenario with limited channel memory, the optimum solution to the combinatorial quadratic optimization problem can be obtained in polynomial time using the classic Viterbi maximum-likelihood sequence estimator [6]. This is useful since it allows us to investigate the quality of the relaxations for larger problem dimensions with moderate computational efforts.

4.1. Signal model

We consider the discrete-time signal model

$$y(t) = H(t)x(t) + v(t) \quad (11)$$

where $y \in \mathbf{C}^m$ is the observed signal, $H \in \mathbf{C}^{m \times n}$ is the system transfer matrix, $x \in \mathbf{C}^n$ is the transmitted signal and $v \in \mathbf{C}^m$ is an unknown noise or perturbation vector. We assume that x is an M -PSK signal with $x_i^M = 1$ where M is a power of 2. We make no assumptions on the structure of H and we assume that v is circularly symmetric zero-mean additive white Gaussian noise with covariance $\sigma^2 I$.

In maximum-likelihood estimation we estimate $x(t)$ given the likelihood function of $y(t)$. For ease of notation we drop the time-dependency in the following. For the simple signal model (11) the likelihood function takes the form

$$f(y) = \frac{1}{\sigma^2 \pi^{nm}} \exp \left\{ -\frac{1}{\sigma^2} \|y - Hx\|^2 \right\}.$$

The maximum likelihood estimation problem is then easily recast as an equivalent minimization problem

$$\begin{aligned} & \text{minimize} && x^* H^* H x - 2 \text{Re } y^* H x \\ & \text{subject to} && x_i^M = 1, \quad i = 1, \dots, n. \end{aligned} \quad (12)$$

For $A = H^* H$ and $b = H^* y$ we have a similar problem as (6).

4.2. Simulation results

We consider a single-user QPSK communication scenario with $x_i^4 = 1$ and a typical channel model with a 5 taps, where each tap is Rayleigh distributed with exponentially decaying powers. All channel realizations are normalized to unit energy and we consider a data-burst length of 100 symbols.

4.2.1. Tightness of the semidefinite relaxation

We start by investigating the tightness of the semidefinite relaxation. Let $q(x) = x^* A x - 2 \text{Re } b^* x$ denote the primal objective and $q(X)$ denote the objective of the relaxations as a function of X , i.e. the objective of either (7) or (9). We define the relative error in percent between the optimal quadratic objective $q(x^*)$ and the semidefinite relaxation $q(X^*)$ as

$$\epsilon(x^*, X^*) = 100 \frac{q(x^*) - q(X^*)}{q(x^*)}.$$

In Fig. 1a we plot the relative error between the optimal solution (obtained by the Viterbi algorithm) and standard semidefinite relaxation of the affine transformation (9) and in Fig. 1b we plot the relative error between the optimal solution and the polynomial relaxation (7). We see that the polynomial relaxation is looser which supports our claim from Sec. 3.

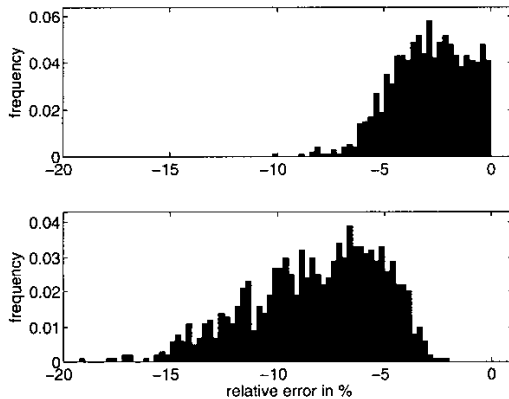


Fig. 1. Tightness of relaxations for the constraint $x_i^4 = 1$ using the two relaxations in Sec. 3 for a fixed SNR=7dB averaged over different channel realizations. Top: relative error of affine transformation, Bottom: relative error of the polynomial relaxation.

4.2.2. Bit-error-probability test

We next investigate the average bit-error-probability achieved with the maximum-likelihood approximation. We use the same channel model described above where a different channel realization is chosen for each data-burst of 100 symbols and assumed to be time-invariant for the duration of a single data-burst. However, we restrict our attention to the boolean case $x_i^2 = 1$ instead of the quaternary case $x_i^4 = 1$. We do this to test the original rounding procedure of Sec. 2. Also, the superior relaxation based on the affine transformation in Sec. 3.2 is easily rewritten as a standard boolean problem; this is not shown here due to space limitations. The results on the quality of the bounds are thus valid for the relaxation in Sec. 3.2. Fig. 1a shows that the .429 bound by Nestorov is in fact quite conservative in practice.

Fig. 2 shows the simulated bit-error-probabilities. For reference we also plot the lowest bound achievable with uncoded BPSK modulation in the absence of ISI.

5. CONCLUSION

In this work we considered extensions of the well-known quadratic problem with boolean variables. The extensions replace the restriction of boolean variables with general polynomial constraints where we considered the polynomial constraint $z^d = 1$.

Recent methods based on semidefinite programming are known to give a provable good bound on the problem solvable in polynomial time using e.g. interior point methods. We considered two different semidefinite relaxations of the problem with higher dimensional constraints. One method applies the standard semidefi-

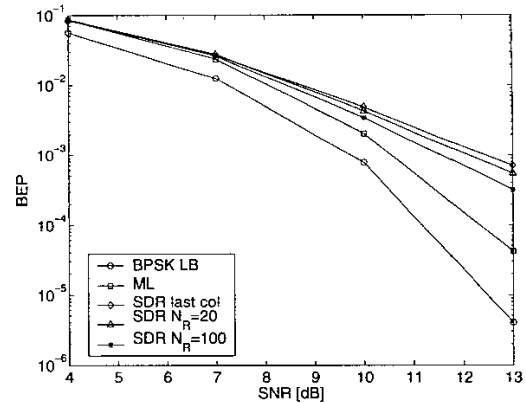


Fig. 2. Simulated bit-error-probabilities. 20000 bursts of 100 bits were demodulated at each SNR. \circ : BPSK lower bound, \square : optimal ML solution, \diamond : thresholded last column of X^* , \triangle : 20 randomization steps, $*$: 100 randomization steps.

nite relaxation to an affine transformation of simpler boolean variables, and another relaxation is based on a semidefinite polynomial relaxation. For the special cases with $z^4 = 1$ we showed that the polynomial relaxation is in fact a further relaxation of the semidefinite relaxation of the affine transformation. Applications of the results include maximum-likelihood estimation in communication systems, where the relaxations offer relatively easily computed approximations of the global optimum. We investigated this application in simulations in order to compare the quality of the different relaxations.

6. REFERENCES

- [1] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1st edition, 1998.
- [2] Michel X. Goemans and David P. Williamson, "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming," *J. Assoc. Comput. Mach.*, vol. 42, pp. 1115–1145, 1995.
- [3] Y. E. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," CORE paper, Louvain-la-Neuve, Belgium, 1997.
- [4] W. K. Ma, T. N. Davidson, K. M. Wong, Z. Q. Luo, and P. C. Ching, "Quasi-maximum-likelihood multiuser detection using semidefinite relaxation with application to synchronous CDMA," *IEEE Trans. on Signal Processing*, vol. 50, no. 4, pp. 912–922, April 2002.
- [5] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE JSAC*, vol. 19, no. 8, pp. 1442–1449, August 2001.
- [6] J. G. Proakis, *Digital Communications*, McGraw Hill, 3rd edition, 1995.
- [7] S. Boyd and L. Vandenberghe, "Convex optimization," Course reader.