Problem structure in semidefinite programs arising in control and signal processing

Lieven Vandenberghe

Electrical Engineering Department, UCLA

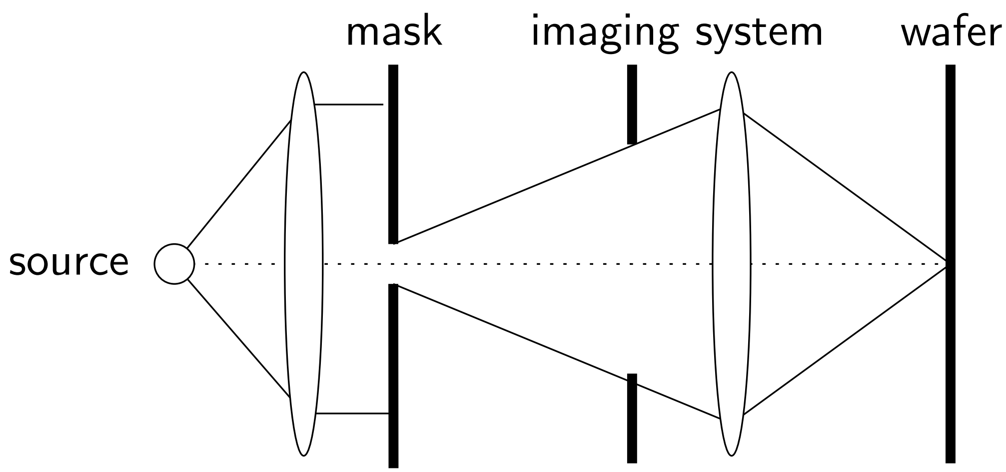
Joint work with: Mehrdad Nouralishahi, Tae Roh

Semidefinite Programming and Its Applications
IMS, National University of Singapore
January 13, 2006

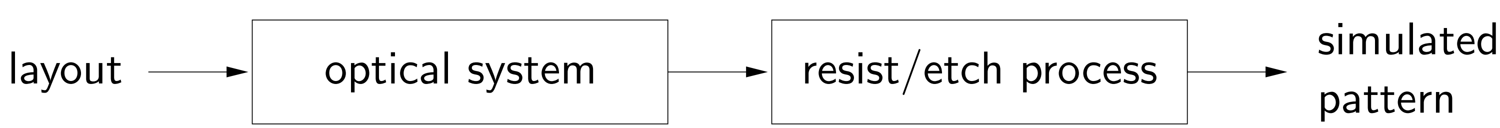
Outline

- **Model calibration in optical lithography.**
- Optimization over nonnegative polynomials.
- Robust least squares.

Model calibration in optical lithography



Model of the lithography process



Model calibration: adjust optical model to include resist/etching effects.

The Hopkins model

$$I(x) = \int \int f(x-v)^* \, K(v)^* \, J(u-v) K(u) \, f(x-u) \, du \, dv$$

$f(x)$: input image (mask); $I(x)$: output image intensity

- Coherent illumination: $J(u-v) = 1$

$$I(x) = \left| \int K(u) f(x-u) du \right|^2 = |(K * f)(x)|^2$$

- Incoherent illumination: $J(u-v) = \delta(u-v)$

$$I(x) = \int |K(u) f(x-u)|^2 \, du$$

- Other choices for J model partially coherent systems.

Optimal coherent approximation

Discrete Hopkins model

$$I(x) = f_x^H W f_x$$

f_x is vector of input values around at x; for example, for 1-D systems,

$$f_x = (f(x-N), f(x-N-1), \ldots, f(x+N))$$

Low-rank approximation

$$I(x) = f_x^H \left(\sum_{k=1}^{r} \phi_k \phi_k^H \right) f_x = \sum_{k=1}^{r} |(\phi_k * f)(x)|^2$$

ϕ_1, . . . , ϕ_r are first few eigenvectors of W.

SDP formulation (for real data)

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(DW) \\ \text{subject to} & -\epsilon I \preceq W - W_0 \preceq \epsilon I \\ & l_k \leq a_k^T W a_k \leq u_k, \quad k = 1, \ldots, m \\ & \text{symmetry constraints on } W \\ & W \succeq 0 \end{array}$$

- W_0 is prior model of optical system.
- Measurements give m upper/lower bounds on intensity for given inputs.
- Objective rewards smoothness or low rank ($D = I$).

Symmetry constraints

One-dimensional system: $W \in \mathbf{S}^n$ is persymmetric, *i.e.*,

$$W = JWJ, \qquad J = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}$$

Two-dimensional system: $W \in \mathbf{S}^{n^2}$ with

$$W = (J_n \otimes I_n)W(J_n \otimes I_n) = (I_n \otimes J_n)W(I_n \otimes J_n)$$

General constraint:

$$W = PWP$$

P a symmetric permutation matrix

Exploiting symmetry

Suppose P is a symmetric permutation matrix ($P^2 = I$):

$$P = \begin{bmatrix} V_\mathrm{e} & V_\mathrm{o} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} V_\mathrm{e} & V_\mathrm{o} \end{bmatrix}^T$$

$V_{\mathrm{e}/\mathrm{o}}$ are eigenvectors with eigenvalue ± 1.

Property: if $W = PWP$, then

$$W = \begin{bmatrix} V_\mathrm{e} & V_\mathrm{o} \end{bmatrix} \begin{bmatrix} W_\mathrm{e} & 0 \\ 0 & W_\mathrm{o} \end{bmatrix} \begin{bmatrix} V_\mathrm{e} & V_\mathrm{o} \end{bmatrix}^T$$

Reduces the number of variables:

- One-dimensional problem: 2 matrices of size $\approx n/2$
- Two-dimensional problem: 4 matrices of size $\approx n^2/4$

Variable upper bounds

LP with upper bounds

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & 0 \preceq x \preceq \mathbf{1} \end{array}$$

- Can introduce slack variables to convert to standard form

$$x + s = \mathbf{1}, \qquad s \succeq 0.$$

  n new variables, n (sparse) equality constraints.

- Upper bounds are easily incorporated in IP methods (at no extra cost).

SDP with upper bounds

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathbf{tr}(A_iX) = b_i, \quad i = 1, \ldots, m \\ & 0 \preceq X \preceq I \end{array}$$

- Introduce slack variables:

$$X + S = I, \qquad S \succeq 0$$

  Adds $n(n+1)/2$ variables and constraints.

- Hence, it is important to handle upper bounds directly.

Newton equations for SDP with upper bounds

$$
\begin{aligned}
-T_1^{-1}\Delta X T_1^{-1} - T_2^{-1}\Delta X T_2^{-1} + \sum_{j=1}^{m} \Delta y_j A_j &= D \\
\mathbf{tr}(A_i \Delta X) &= d_i, \quad i = 1, \ldots, m
\end{aligned}
$$

where $T_1, T_2 \succ 0$.

- Find congruence that simultaneously diagonalizes T_1 and T_2:

$$R^T T_1 R = I, \qquad R^T T_2 R = \mathbf{diag}(\gamma)^{-1}$$

- Make change of variables $\Delta \tilde{X} = R^T \Delta X R$.

- Transformed equations:

$$\begin{aligned} -\Delta\tilde{X} - \mathbf{diag}(\gamma)\Delta\tilde{X}\,\mathbf{diag}(\gamma) + \sum_{j=1}^{m} \Delta y_j \tilde{A}_j &= \tilde{D} \\ \mathbf{tr}(\tilde{A}_i \Delta\tilde{X}) &= d_i, \quad i = 1, \ldots, m \end{aligned}$$

- Eliminate $\Delta\tilde{X}$ from first equation:

$$\Delta\tilde{X} = \sum_{j=1}^{m} \Delta y_j (\tilde{A}_j \circ \Gamma) - \tilde{D} \circ \Gamma, \qquad \Gamma_{kl} = 1/(1 + \gamma_k \gamma_l)$$

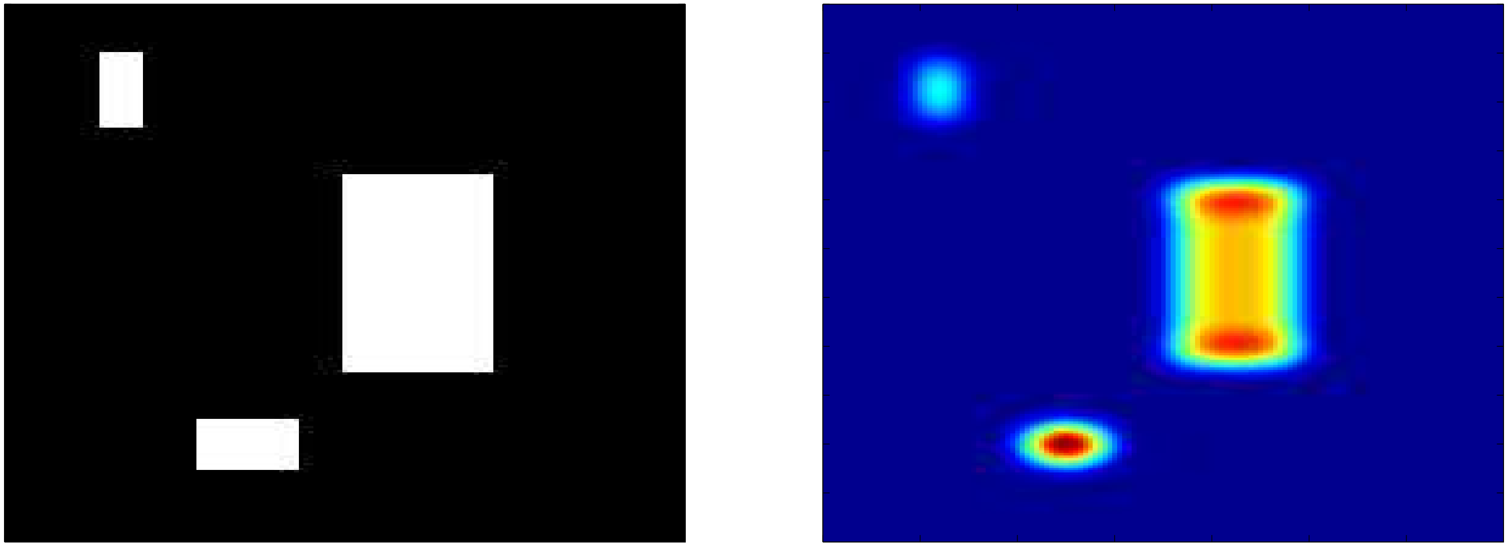
- Solve

$$\sum_{j=1}^{m} \mathbf{tr}(\tilde{A}_i(\tilde{A}_j \circ \Gamma))\Delta y_j = b_i, \quad i = 1, \ldots, m$$

Cost: comparable to SDP without upper bounds.

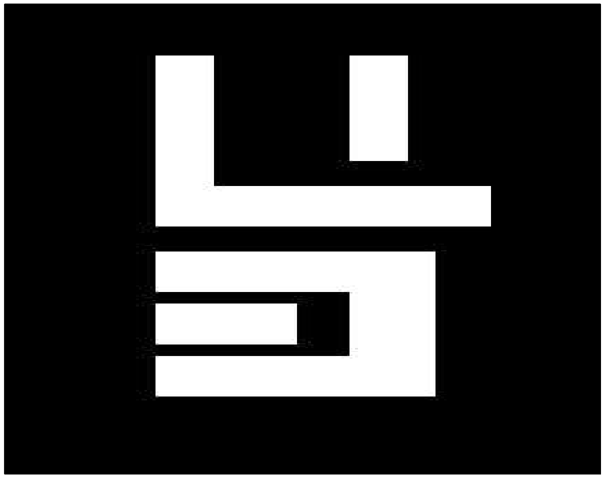
Example

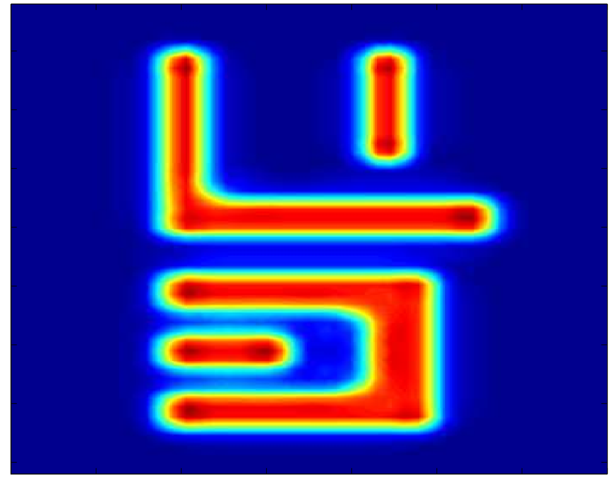
Calibration step (using simulated data)



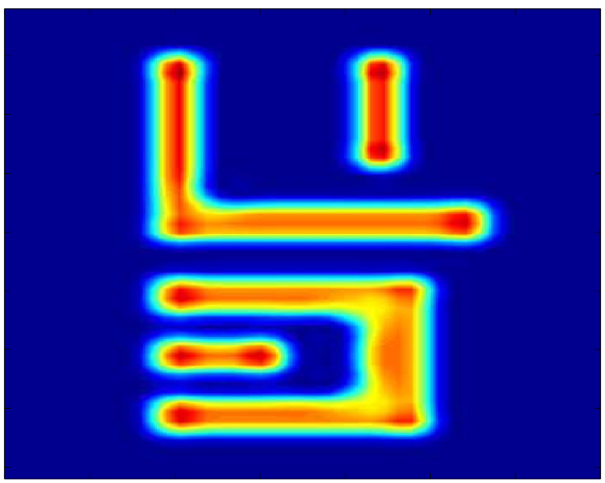
- Left: mask. Right: measured intensity.
- W, $W_0 \in \mathbf{S}^{1849}$.
- Sample output image at $m = 2500$ points.

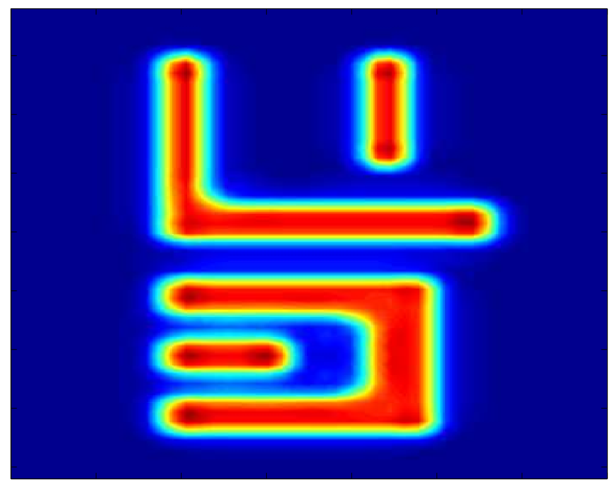
Test input and actual output





Predicted output with uncalibrated and calibrated model





Outline

- Model calibration in optical lithography.
- **Optimization over nonnegative polynomials.**
- Robust least squares.

A useful SDP standard form

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(BX) \\ \text{subject to} & A\,\mathbf{diag}(CXC^T) = b \\ & X \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & C^T\,\mathbf{diag}(A^T y)C \preceq B \end{array}$$

Interpretation (for $A \in \mathbf{R}^{m\times r}$ with rows a_i^T, $C \in \mathbf{R}^{r\times n}$)

- Constraints are equivalent to $\mathbf{tr}(F_i X) = b_i$, $i = 1, \ldots, m$, where

$$F_i = C^T\,\mathbf{diag}(a_i)C, \quad i = 1, \ldots, m.$$

  Matrices F_i are linear combinations of r rank-one matrices.

- Possible for arbitrary F_i if $r = mn$. Interesting when $r \ll mn$.
- F_i's can be **dense** and **full-rank**.

Newton equations

$$
\begin{aligned}
-T^{-1}\Delta X T^{-1} + C^T \,\mathbf{diag}(A^T \Delta y) C &= D \\
A \,\mathbf{diag}(C \Delta X C^T) &= d
\end{aligned}
$$

Eliminate ΔX from first equation, substitute in second: $H\Delta y = g$ with

$$H = A\left((CTC^T) \circ (CTC^T)\right) A^T$$

Cost if $m = O(n)$, $r = O(n)$: $O(n^3)$ operations

SDP formulation of sums of squares

$$x^T f(t) = \sum_{k=1}^{s} (y_k^T q(t))^2$$

- r.h.s. sampled at $t_1, \ldots, t_N$,

$$\sum_{k=1}^{s} ((Cy_k) \circ (Cy_k)) = \sum_k \mathbf{diag}(Cy_k y_k^T C^T) = \mathbf{diag}(CXC^T)$$

where $C = \begin{bmatrix} q(t_1) & q(t_2) & \cdots & q(t_N) \end{bmatrix}^T$, $X = \sum_k y_k y_k^T$.

- Coefficients of l.h.s. from samples at $t_1, \ldots, t_N$:

$$x = A\,\mathbf{diag}(CXC^T)$$

Applications

- Trigonometric polynomial nonnegative on (subinterval of) $[0, 2\pi]$
(via discrete Fourier transform).

- Cosine polynomials nonnegative on (subinterval of) $[0, \pi]$
(via discrete cosine transform).

- Real polynomials nonnegative on (subintervals of) **R**
(via discrete polynomial transforms and orthogonal polynomials).

- Sum-of-squares representations of multivariate nonnegative polynomials.

Nonnegative trigonometric polynomial

$$x_0 + 2\Re(x_1 e^{-j\omega} + \cdots + x_n e^{-jn\omega}) = \left| y_0 + y_1 e^{-j\omega} + \cdots + y_n e^{-jn\omega} \right|^2$$

- r.h.s. at $\omega = 2\pi k/N$, $k = 0, \ldots, N-1$, where $N \geq 2n+1$:

$$(Wy) \circ (\overline{Wy}) = \mathbf{diag}(WXW^H), \qquad X = yy^H$$

  W: first $n+1$ columns of DFT matrix of length N.

- Inverse DFT maps sampled values to coefficients of l.h.s.:

$$x = \frac{1}{N} W^H \, \mathbf{diag}(WXW^H) \tag{1}$$

Hence, polynomial is nonnegative iff (1) for some $X \succeq 0$.

Dual interpretation: parametrization of Toeplitz matrix

$$\frac{1}{2}\begin{bmatrix} 2z_0 & \bar{z}_1 & \cdots & \bar{z}_n \\ z_1 & 2z_0 & \cdots & \bar{z}_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ z_n & z_{n-1} & \cdots & 2z_0 \end{bmatrix} = \frac{1}{N} W^H \mathbf{diag}(\Re(Wz)) W$$

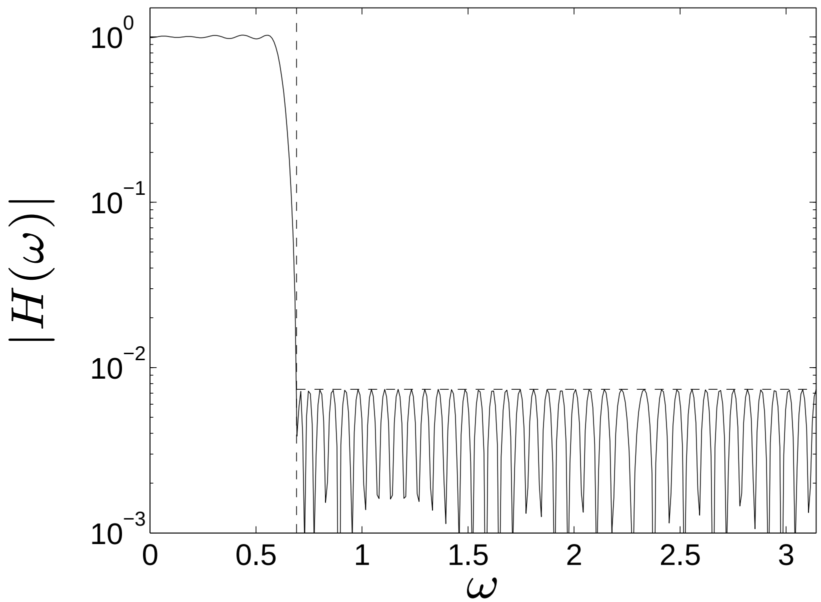
Follows from expressions for convolution of z with a vector u:

$$\begin{aligned} z * u = \begin{bmatrix} z_0 & 0 & \cdots & 0 \\ z_1 & z_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ z_n & z_{n-1} & \cdots & z_0 \end{bmatrix} u &= \frac{1}{N} W^H ((Wz) \circ (Wu)) \\ &= \frac{1}{N} W^H \mathbf{diag}(Wz) W u \end{aligned}$$

Example: Linear-phase Nyquist filter

$$\text{minimize} \quad \sup_{\omega \geq \omega_\mathrm{s}} |h_0 + h_1 \cos\omega + \cdots + h_n \cos n\omega|$$

with $h_0 = 1/M$, $h_{kM} = 0$ for positive integer k.



(Example with $n = 50$, $M = 5$, $\omega_\mathrm{s} = 0.69$.)

SDP formulation

Semi-infinite LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t \leq h_0 + h_1 \cos\omega + \cdots + h_n \cos n\omega \leq t, \qquad \omega_{\rm s} \leq \omega \leq \pi \end{array}$$

SDP formulation

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & h + te_0 = A_1\,\mathbf{diag}(C_1X_1C_1^T) + A_2\,\mathbf{diag}(C_2X_2C_2^T) \\ & -h + te_0 = A_1\,\mathbf{diag}(C_1X_3C_1^T) + A_2\,\mathbf{diag}(C_2X_4C_2^T) \\ & X_1 \succeq 0, \quad X_2 \succeq 0, \quad X_3 \succeq 0, \quad X_4 \succeq 0 \end{array}$$

- Variables: t, h_i (except for $i = kM$), 4 matrices X_i of size roughly $n/2$.
- A_i, C_i constructed from DCT matrices of length $N \geq n$.

Nonnegative polynomial on R

$$\begin{aligned} x_0p_0(t) + x_1p_1(t) + \cdots + x_np_n(t) \\ = \quad (u_0p_0(t) + \cdots + u_mp_m(t))^2 + (v_0p_0(t) + \cdots + v_mp_m(t))^2 \end{aligned}$$

(p_k: basis polynomial of degree k; $n = 2m$)

- r.h.s. evaluated at t_0, . . . , t_N:

  $$(Cu) \circ (Cu) + (Cv) \circ (Cv) = \mathbf{diag}(CXC^T), \qquad X = uu^T + vv^T$$

  where $C_{ij} = p_j(t_i)$

- If $N \geq n$ and t_i are distinct, the sample values uniquely specify x:

  $$x = A\,\mathbf{diag}(CXC^T) \qquad (2)$$

Polynomial is nonnegative iff (2) holds for some $X \succeq 0$.

Discrete polynomial transform

$$V_{\mathrm{DPT}} = \begin{bmatrix} p_0(t_0) & p_1(t_0) & \cdots & p_N(t_0) \\ p_0(t_1) & p_1(t_1) & \cdots & p_N(t_1) \\ \vdots & \vdots & & \vdots \\ p_0(t_N) & p_1(t_N) & \cdots & p_N(t_N) \end{bmatrix}, \qquad W_{\mathrm{DPT}} = V_{\mathrm{DPT}}^{-1}$$

- p_0, p_1, . . . : system of orthogonal polynomials
- t_0, . . . , t_N: roots of p_{N+1}
- V_{DPT} maps coefficients of

  $$x_0p_0(t) + x_1p_1(t) + \cdots + x_Np_N(t)$$

  to values at t_0, . . . t_N
- W_{DPT} maps values at t_0, . . . , t_N to coefficients

Three-term recursion

$$t\begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_N(t) \end{bmatrix} = \begin{bmatrix} \beta_0 & \alpha_0 & 0 & \cdots & 0 \\ \alpha_0 & \beta_1 & \alpha_1 & \cdots & 0 \\ 0 & \alpha_1 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_N \end{bmatrix} \begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \\ p_N(t) \end{bmatrix} + \alpha_N \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ p_{N+1}(t) \end{bmatrix}$$

$$tp(t) = Jp(t) + \alpha_N p_{N+1}(t) e_N$$

- Eigenvalues of J are the roots of p_{N+1}:

$$t_i p(t_i) = Jp(t_i)$$

- V_{DPT}, W_{DPT} easily computed from eigenvalue decomposition of J.

Numerical example

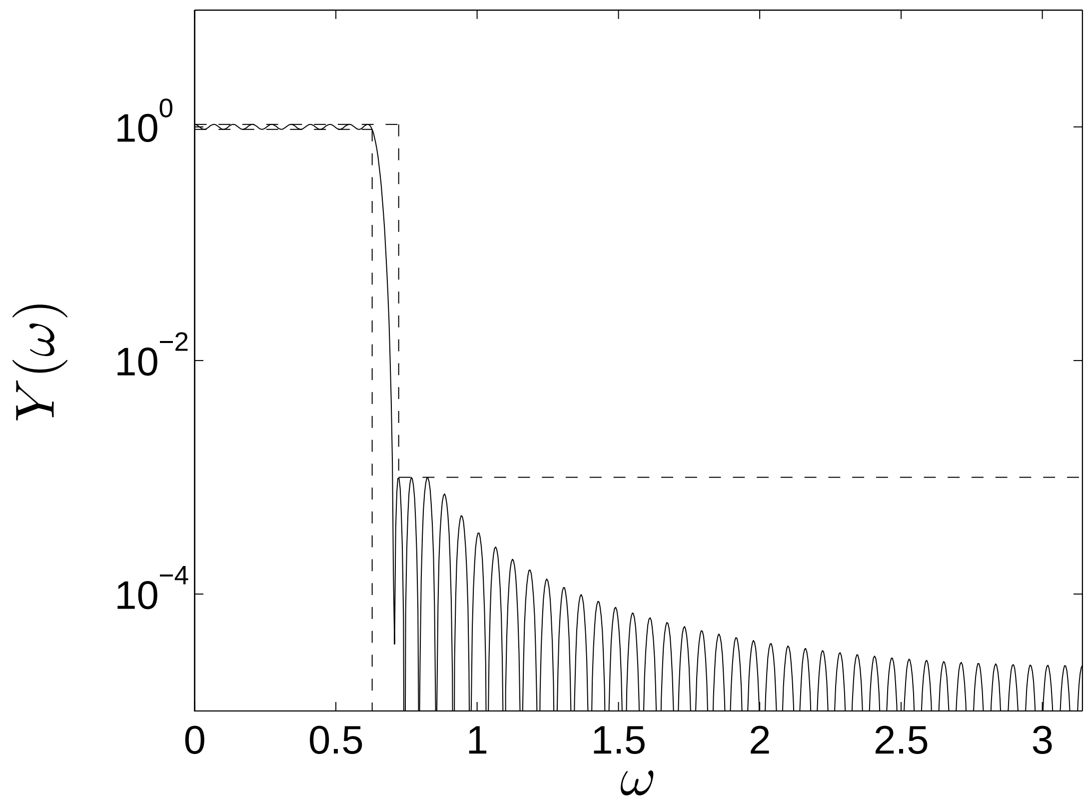
Magnitude FIR filter design

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\int_{\omega_{\rm s}}^{\pi} Y(\omega)\, d\omega \\
\text{subject to} & 1/\delta_{\rm p} \leq Y(\omega) \leq \delta_{\rm p}, \quad 0 \leq \omega \leq \omega_{\rm p} \\
& Y(\omega) \leq \delta_{\rm s}, \quad \omega_{\rm s} \leq \omega \leq \pi \\
& Y(\omega) \geq 0, \quad 0 \leq \omega \leq \pi
\end{array}
$$

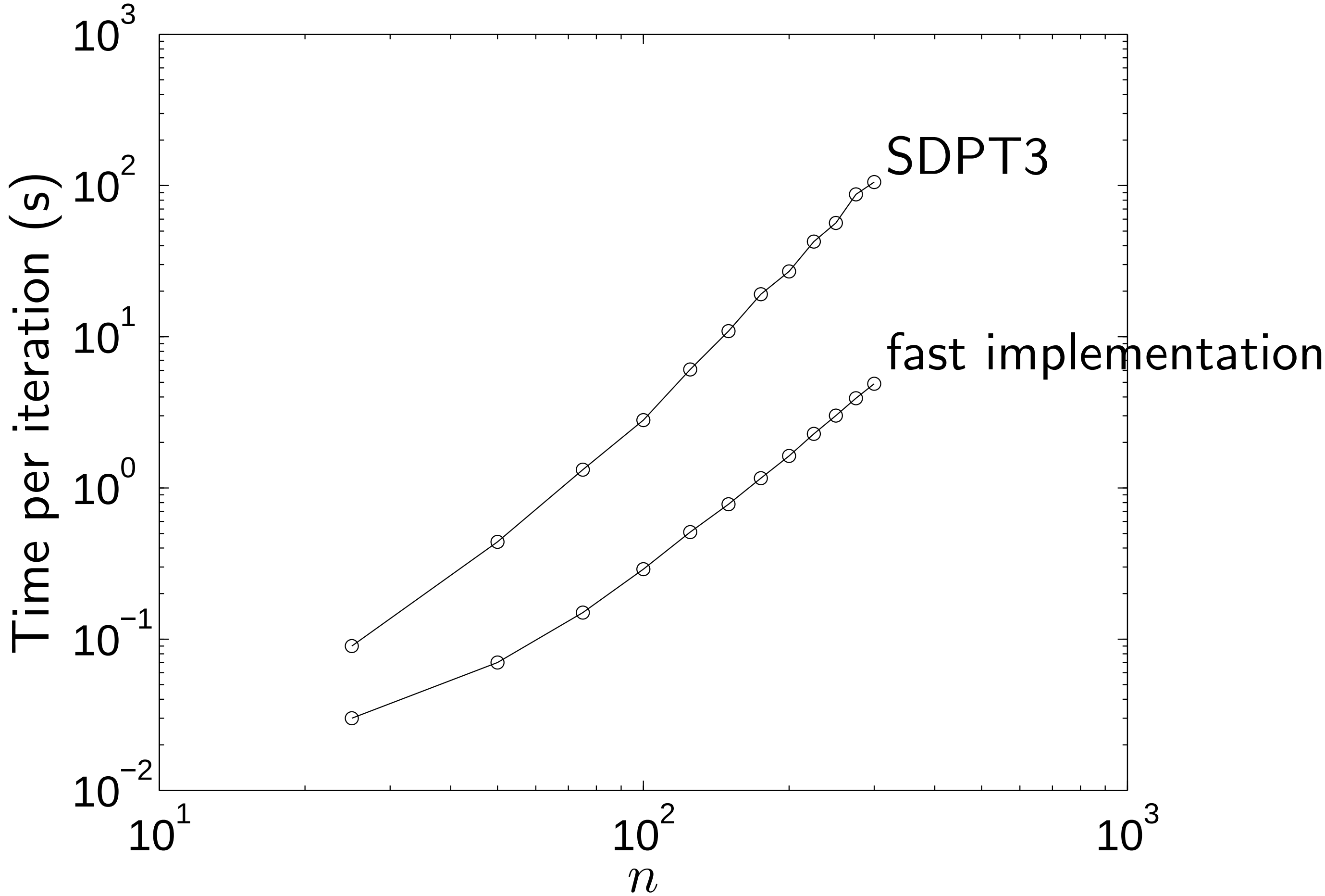
where $Y(\omega) = y_0 + y_1 \cos\omega + \cdots + y_n \cos n\omega$.

- Constraints result in 4 LMI constraints.
- Variables: y and 8 auxiliary matrix variables of size roughly $n/2$.

Example $(n = 101)$



Time per iteration (Matlab on 2.8GHz P4)



SDPT3-style primal-dual method with fast solution of Newton equations

Outline

- Model calibration in optical lithography.
- Optimization over nonnegative polynomials.
- **Robust least squares.**

Robust least-squares

$$\text{minimize} \quad \sup_{\|u\|_2 \leq 1} \|A(u)x - b\|_2^2$$

where

$$A(u) = \bar{A} + U\,\mathbf{diag}(Du)V^T$$

Example. A is lower triangular Toeplitz with coefficients $h_k + u_k$

$$A(u) = \frac{1}{N} W_1^H\,\mathbf{diag}(W_1(h+u))W_2$$

W_1, W_2: first $n+1$, resp. $m+1$, columns of DFT matrix

SDP formulation

$$\begin{array}{ll} \text{minimize} & t + \lambda \\ \text{subject to} & \begin{bmatrix} t & 0 & (\bar{A}x - b)^T \\ 0 & \lambda I & D^T \,\mathbf{diag}(V^T x) U^T \\ (\bar{A}x - b) & U \,\mathbf{diag}(V^T x) D & 0 \end{bmatrix} \succeq 0 \end{array}$$

Cost per iteration is dominated by constructing the matrix

$$V \left((DT_{23}U) \circ (DT_{23}U)^T + (DT_{22}D^T) \circ (U^T T_{33} U) \right) V^T$$

T_{ij} are submatrices of scaling matrix

$O(n^3)$ operations (if all dimensions are of the same order)

Conclusions

Some interesting types of problem structure that are easily exploited.

- Symmetry constraints.
- Upper bounds on the matrix variables.
- A generalization of low-rank structure.