Convex and Nonsmooth Optimization
Problem Sets w/ Solutions

Rex C. Liu, Courant Institute of Mathematical Sciences

October 3, 2025

Disclaimer:

These are the problem sets for the course *Convex and Nonsmooth Optimization*, given by Dr. Michael L. Overton, Silver Professor of Computer Science and Mathematics, at New York University in Spring 2024.

Convex optimization problems have many important properties, including a powerful duality theory and the property that any local minimum is also a global minimum. Nonsmooth optimization refers to minimization of functions that are not necessarily convex, usually locally Lipschitz, and typically not differentiable at their minimizers. Topics in convex optimization that will be covered include duality, linear and semidefinite programming, CVX ("disciplined convex programming"), gradient and Newton methods, Nesterov's lower complexity bound and optimal gradient method, the alternating direction method of multipliers, the nuclear norm and matrix completion, primal-dual interior-point methods for linear and semidefinite programs. Topics in nonsmooth optimization that will be covered include subgradients and subdifferentials, Clarke regularity, and algorithms, including gradient sampling, BFGS and the stochastic gradient method, for nonsmooth, nonconvex optimization. The text is [BV04]. Other references include [Nes18].

The solutions are mostly given by Rex Liu with help from Tao Li, Yizheng (Thomas) Li, Letao (Jenna) Chen, and Zhen (Bobby) Bao. If you see any mistakes or think that the presentation is unclear and could be improved, please send an email to: cl5682@nyu.edu. All comments and suggestions are appreciated.

Notations:

- $\mathbb{S}_+^n$: the set of symmetric positive semidefinite $n \times n$ matrices.
- $\mathbb{S}_{++}^n$: the set of symmetric positive definite $n \times n$ matrices.
- K^*: the dual cone of the cone K.
- $x \succeq y, y \preceq x$: componentwise inequality between vectors x and y.
- $\mathbf{1}$: vector with all components one.
- $\mathrm{diag}(x)$: diagonal matrix with diagonal entries $x_1, \ldots, x_n$.
- $\mathrm{Conv}(C)$: the convex hull of a set C.

Contents

```
    cvx_end

    % Generate the figure
    figure;
    x = linspace(-10, 10); % Adjust this range as needed
    theta = 0:pi/100:2*pi;
    hold on
    for k = 1:colsA
        if A(2, k) ~= 0
            plot(x, -x*A(1, k)./A(2, k) + b(k)./A(2, k), 'b-');
        else
            y_line = b(k) / A(1, k); % For horizontal lines
            plot([min(x), max(x)], [y_line, y_line], 'b-');
        end
    end
    plot(x_c(1) + r*cos(theta), x_c(2) + r*sin(theta), 'r');
    plot(x_c(1), x_c(2), 'k+');
    xlabel('x_1');
    ylabel('x_2');
    title('Largest Euclidean ball lying in a 2D polyhedron');
    axis equal;
    hold off
end
```

Figure 1 shows the example on the webpage (left) as well as another example with more inequalities (right). ☐

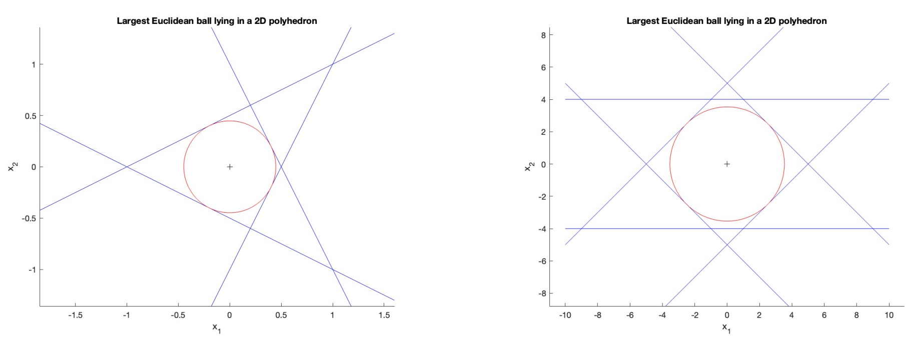


Figure 1: Largest Euclidean ball lying in a 2D polyhedron

7. What happens if you choose A and b so there is no point inside the polyhedron?

Proof. It will be an infeasible region (an empty polyhedron). In MATLAB when running such

```
    title(['Largest ball in ' num2str(p) '-norm']);
end
```

Here are the plots (Figure 2) for $p = 1, 1.5, \infty$, respectively. $\square$

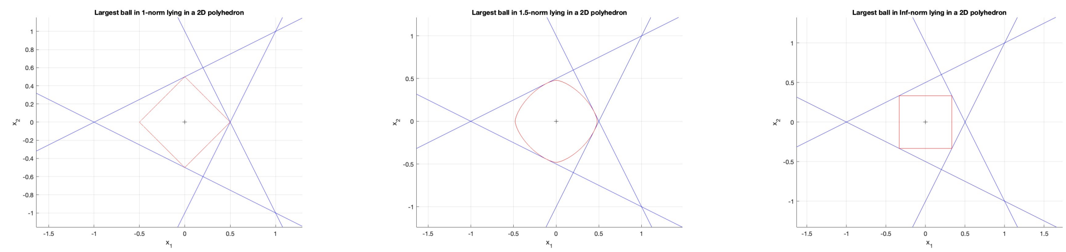


Figure 2: Largest p-norm ball

9. Explain why your answer to the previous question does not extend to work for $p < 1$, for which MatLab's `norm(x,p)` is still well defined? What goes wrong if you try it anyway?

   Solution. It is because when $p < 1$ it does not satisfy the triangle inequality, thus not a norm by definition. If we try it anyway, it does not fit as the correct "largest ball." $\square$

10. Is there a simple way to change the code so that it will actually still give the right answer when $p < 1$? Justify your answer and run and show the output for an example.

    Solution. A simple way is to change the eighth line of code p==1 into p<=1. The reason is that the convex polyhedron cannot intersect the concave regions of a non-convex ball. An example is given in Figure 3 for $p = 1/2$. $\square$

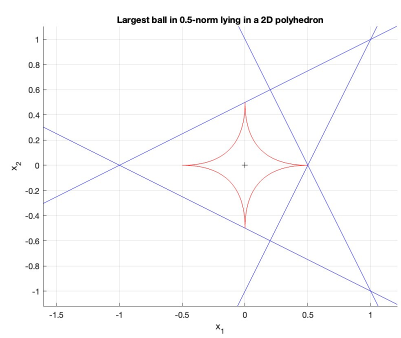


Figure 3: Largest ball for $p < 1$

Solution. Note that $x \geq 0 \iff -x \leq 0$. By definition, the Lagrangian is

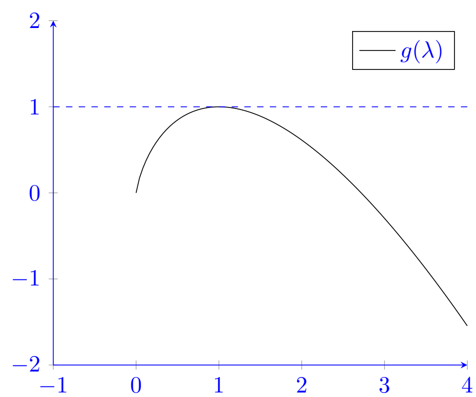
$$L(x, \lambda) = e^x - \lambda x$$

The Lagrange dual is obtained by minimizing the Lagrangian over x, for a given λ:

$$g(\lambda) = \inf_{x \geq 0} L(x, \lambda) = \inf_{x \geq 0} (e^x - \lambda x)$$

Taking the derivative while checking the valid domain of λ, one gets

$$g(\lambda) = \begin{cases} -\infty & \lambda < 0 \\ \lambda - \lambda \ln(\lambda) & \lambda \geq 0 \end{cases}$$



The plot is sketched above. □

(b) Solve the Lagrange dual problem by maximizing $g(\lambda)$ over $\lambda \geq 0$. Does strong duality hold?

Proof. The dual problem is to

$$\begin{array}{ll} \text{maximize} & \lambda - \lambda \ln(\lambda) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

By taking the derivative of $g(\lambda)$, we see that $\sup g(\lambda)$ is reached at $g(1) = 1$. The dual problem has an optimal value $d^* = 1$. Since $p^* = d^*$, strong duality holds. □

3. Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & x^2 + 1 \\ \text{subject to} & (x-2)(x-4) \leq 0 \end{array}$$

with variable $x \in \mathbb{R}$.

(a) Give the feasible set, the optimal value, and the optimal solution.

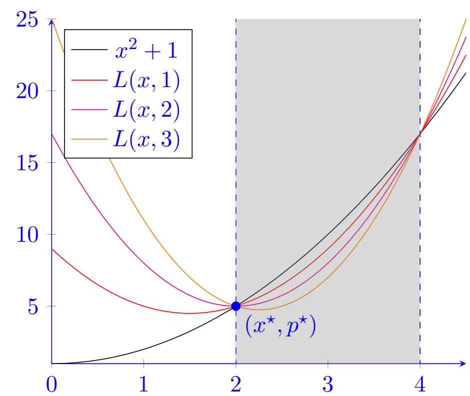
Solution. Solving $(x-2)(x-4) \leq 0$ yields $x \in [2,4]$, which is the feasible set. Since x^2+1 is monotonically increasing in the feasible set, the optimal value is reached at $x^\star = 2$ with $p^\star = 2^2 + 1 = 5$. □

(b) Plot the objective x^2+1 versus x. On the same plot, show the feasible set, optimal point and value, and plot the Lagrangian $L(x,\lambda)$ versus x for a few positive values of λ. Verify the lower bound property ($p^\star \geq \inf_x L(x,\lambda)$ for $\lambda \geq 0$). Derive and sketch the Lagrange dual function g.

Solution. By definition, the Lagrangian is

$$L(x,\lambda) = (1+\lambda)x^2 - 6\lambda x + (1+8\lambda)$$

The plot of the objective x^2+1 as well as $L(x,\lambda)$, $\lambda = 1,2,3$ versus x is given as follows:



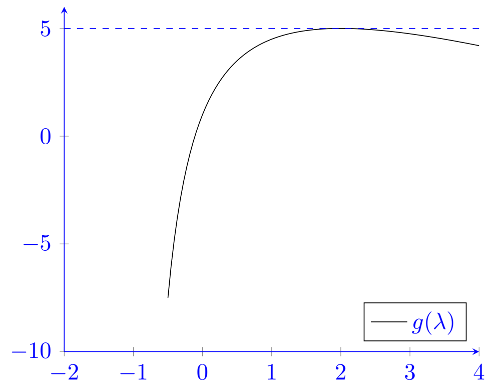
The lower bound property is verified that the minimum value of $L(x,\lambda)$ over x is always less than $p^\star$. Note that for $\lambda > 0$, by the AM-GM inequality,

$$\inf_x L(x,\lambda) = 1 + 8\lambda - \frac{9\lambda^2}{1+\lambda} \overset{k=1+\lambda}{=} 11 - k - \frac{9}{k} \leq 5 = p^\star$$

where equiality is reached at $k=3$, i.e. $\lambda = 2$.

Note that the above infimum is valid for $\lambda > -1$, since f_0, f_1 both have principal coefficient 1. For $\lambda \leq -1$ it is unbounded below. Thus

$$g(\lambda) = \begin{cases} 1 + 8\lambda - \frac{9\lambda^2}{1+\lambda} & \lambda > -1 \\ -\infty & \lambda \leq -1 \end{cases}$$



The plot is sketched above. $\square$

(c) State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution $\lambda^\star$. Does strong duality hold?

Solution. The Lagrange dual problem is

$$\begin{array}{ll} \text{maximize} & 1+8\lambda-\frac{9\lambda^2}{1+\lambda} \\ \text{subject to} & \lambda \geq 0 \end{array}$$

It is a concave maximization problem as we have checked that the infimum is reached at $\lambda = 2$ and $g(\lambda)$ is decreasing either when λ goes down to 0 or goes up to ∞. The dual optimal value is reached at $\lambda^\star = 2$ with $d^\star = 5$. The strong duality holds as $d^\star = p^\star$. $\square$

4. *A convex problem in which strong duality fails.* Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq 0 \end{array}$$

with variables x and y, and domain $\mathcal{D} = \{(x,y) \mid y > 0\}$.

(a) Verify that this is a convex optimization problem. Find the optimal value.

Solution. For e^{-x}, note that $(e^{-x})'' = e^{-}x > 0$ for all x, so $f_0(x)$ is convex. For x^2/y, note that the Hessian matrix

$$\begin{pmatrix} \frac{2}{y} & \frac{-2x}{y^2} \\ \frac{-2x}{y^2} & \frac{2x^2}{y^3} \end{pmatrix}$$

is positive semidefinite for $y > 0$ since its eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 2(x^2+y^2)/y^3$, so $f_1(x,y)$ is convex. Therefore, it is indeed a convex optimization problem.

Note that $x^2 \geq 0$ for all x but the domain restricts $y > 0$, and the inequality constraint further restricts $x^2 \leq 0$, implying that x can only be 0, so the optimal value $p^\star = 1$. $\square$

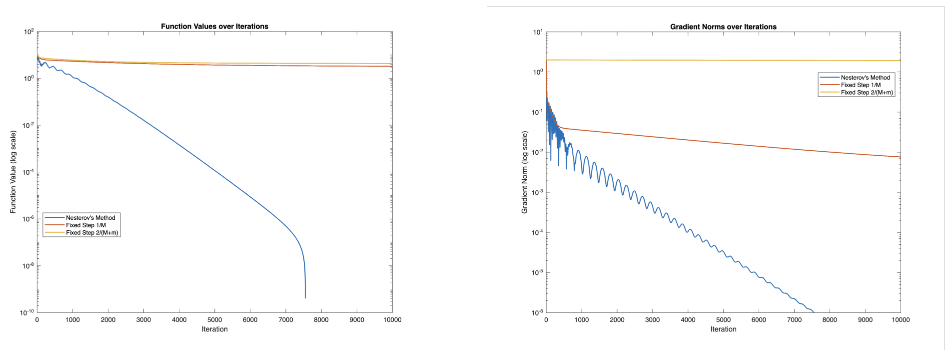


Figure 7: Comparison of three gradient methods for example 1

decreases much slower than $O(1/k)$ for $t = 1/M$ when the number k of iterations is not large enough. It is thus valid in the plot that gradient with step size $1/M$ converges faster.

For the second example, I chose $M = 564$ and $m = 2$, which were estimated in the previous section. See Figure 8 for the result, which aligns with the theoretical guarantees.

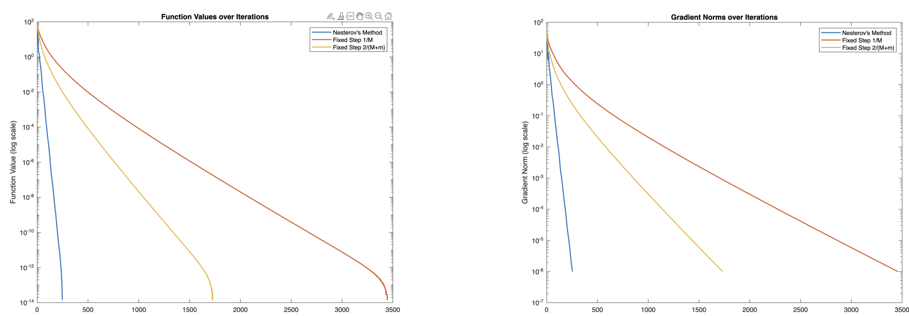


Figure 8: Comparison of three gradient methods for example 2

For the last two examples, we first code matrix `T` (see [Nes18, p. 78, matrix A]) where its diagonal has entries 2 while sub-diagonals have entries -1 *in an efficient way*:

```
T = 2*speye(dim); % Main diagonal
T(2:dim+1:end) = -1; % upper subdiagnoal
T(dim+1:dim+1:end) = -1; % lower subdiagnoal
```

Then we code the function the calculate f and ∇f:

```
function [f, g] = badFunction(x, m, M, T, e1)
% Nesterove's worst case example
% input: m, M: mu and L in book; T: matrix A in book; e1: e_1 in book

% Calculate the quadratic form component of the function value
quadraticForm = norm(x(1:end-1) - x(2:end))^2;
% Incorporate the first element's squared value and adjustment
quadraticForm = x(1)^2 + quadraticForm - 2*x(1);
% Calculate f
f = (M - m) / 8 * quadraticForm + m / 2 * norm(x)^2;

% Compute the gradient 'g'
g = (M - m) / 4 * T * x + m * x - (M - m) / 4 * e1;
end
```

See Figure 9 and 10, respectively, for the results. It is checked that $mI \preceq \nabla^2 F \preceq MI$ by

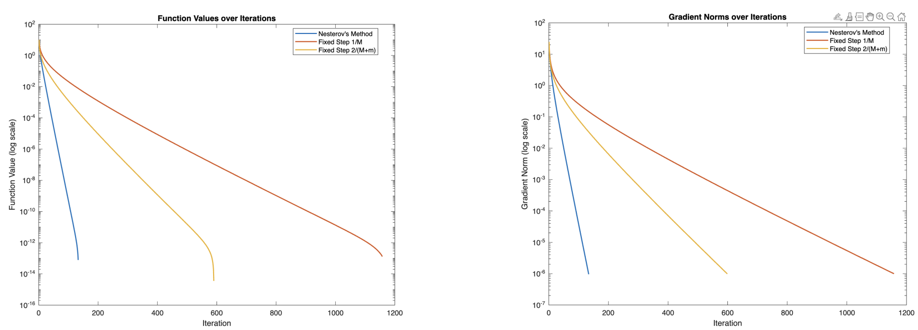


Figure 9: Comparison of three gradient methods for example 3

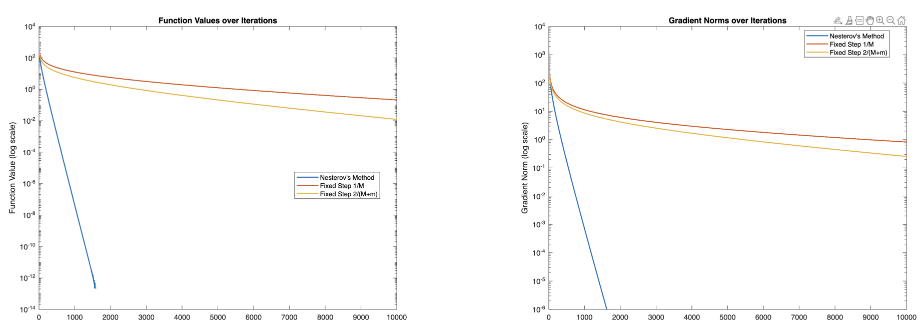


Figure 10: Comparison of three gradient methods for example 4

$\rho \backslash \lambda$	0.01	0.1	1
0.1	8586	8298	7864
1	8519	7870	7054
10	8551	7927	4932

Table 1: Large problem numbers of nonzeros for different ρ and λ combinations

```
spy(L);
title('Sparsity Pattern of Cholesky Factor L');
```

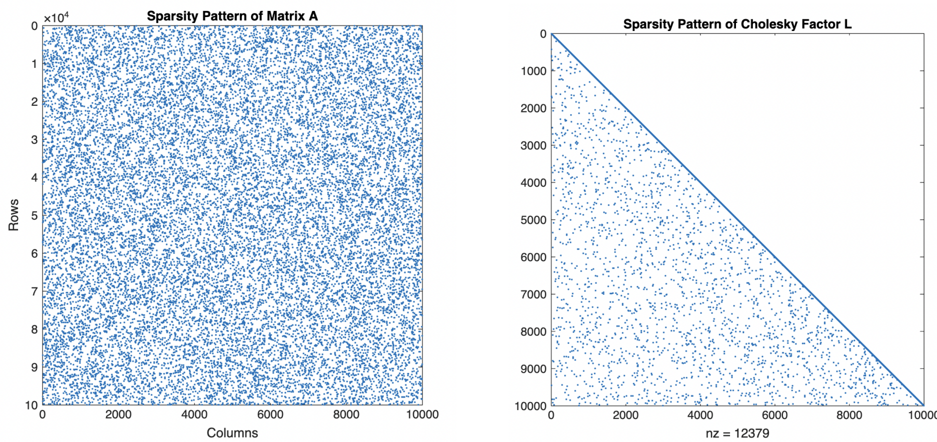


Figure 11: Sparsity patterns of randomly generated matrices A and L, with ρ picked as 1

□

Then, inside the ADMM iteration you can solve the relevant systems with forward and back substitution [BV04, Algorithm C.2, p.670], using the backslash operator \.

For both the small problem and the large problem:

- Experiment with λ : does larger λ result in solutions x which are more sparse, as it should?
- Experiment with ρ : what effect does this have on the method?

Solution. Please see Figures 12 and 13 for the logplots. Sparsity of the small problem remains unchanged with all same nz values, probably due to its small size. Sparsity of the large problem has the following trend (for a random trial):

The reduction in nonzeros as λ increases confirms that the regularization effectively promotes sparsity in the solution. The effect of varying ρ on the sparsity of the solution is less pronounced

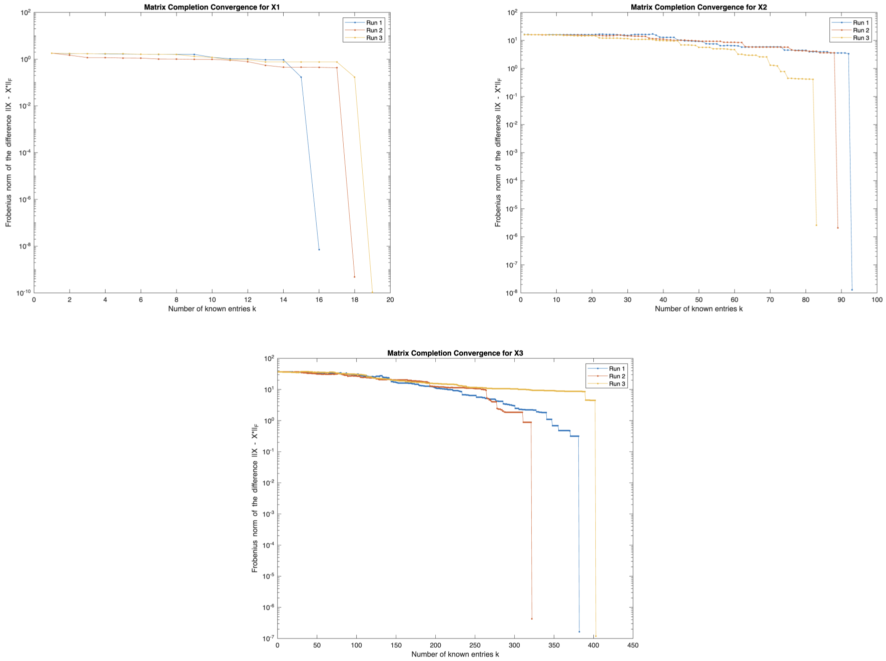


Figure 14: Logplots of Matrix Completion Convergence for X_1, X_2, X_3

Proof. We want to show that

$$\|X\| \leq \|X\|_F \leq \|X\|_* \leq \sqrt{r}\|X\|_F \leq r\|X\|$$

for any matrix X of rank at most r, where σ_i's are the singular values of X, which are all positive, $\|X\|_F = \left(\sum_{i=1}^r \sigma_i^2\right)^{\frac{1}{2}}$, $\|X\| = \sigma_1(X)$, and $\|X\|_* = \sum_{i=1}^r \sigma_i(X)$.

The first inequality $\|X\| \leq \|X\|_F$ is trivially true, since $\sigma_1^2 \leq \sum_i \sigma_i^2$. Then, since the singular values are all positive, we have $\sum_i \sigma_i^2 \leq (\sum_i \sigma_i)^2$, so that $\|X\|_F \leq \|X\|_*$. By the AM-QM inequality[5], we also have $(\sigma_1 + \cdots + \sigma_r)/r \leq \sqrt{\left(\sum_i \sigma_i^2\right)/r}$. Multiplying r on both sides yields $\|X\|_* \leq \sqrt{r}\|X\|_F$. The last inequality $\sqrt{r}\|X\|_F \leq r\|X\|$ is again trivial, as it is equivalent saying, by taking squares on both sides, that $\sigma_1^2 + \cdots + \sigma_r^2 \leq \sigma_1^2 + \cdots + \sigma_1^2$, which is true as σ_1 is the largest of all σ_i. $\square$

[5]The arithmetic mean and quadratic mean inequality states that, for positive reals x_i, we have

$$\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2 \leq \frac{\sum_{i=1}^n x_i^2}{n},$$

which can be proved by applying the Cauchy–Schwarz inequality on vectors $\mathbf{x}$ and $\mathbf{1}$.

5. Consider the nuclear norm relaxation of the generalization of the matrix completion problem stated on [RFP10, p. 480]. Show that the program written on the right side is indeed its dual. Here, if $A(X) = b$ means $\langle A_k, X\rangle = b_k$, $k = 1, \ldots, p$, then $A^*(z)$ means the adjoint operation, $\sum_k z_k A_k$.

   Proof. The primal problem is:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}(\operatorname{tr}(W_1) + \operatorname{tr}(W_2)) \\ \text{subject to} \quad & \begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix} \succeq 0 \\ & \mathcal{A}(X) = b \end{aligned}$$

   Note that W_1 and W_2 are symmetric. The constraint $\mathcal{A}(X) = b$ can be expanded to $\langle A_k, X\rangle = b_k$ for $k = 1, \ldots, p$, where $\{A_k\}$ are given matrices forming the linear operator $\mathcal{A}$.

   Construct the Lagrangian L with dual variables Z (a symmetric matrix) for the semidefinite constraint and $z \in \mathbb{R}^p$ (vector) for the linear constraints:

$$\begin{aligned} & L(W_1, W_2, X, Z, z) \\ &= \frac{1}{2}(\operatorname{tr}(W_1) + \operatorname{tr}(W_2)) - \operatorname{tr}\left(Z \begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix}\right) + \langle z, b - \mathcal{A}(X)\rangle \\ &= \frac{1}{2}\operatorname{tr}(W_1(I - 2Z_{11})) + \frac{1}{2}\operatorname{tr}(W_2(I - 2Z_{22})) \\ &\quad - \operatorname{tr}(Z_{12}X) - \operatorname{tr}(Z_{21}X^\top) - \operatorname{tr}(\mathcal{A}^*(z)X) + b^\top z \end{aligned}$$

   where $Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$.

   The dual function $g(z, Z)$ is obtained by minimizing L with respect to W_1, W_2, and X:

$$g(z, Z) = \inf_{W_1, W_2, X} L$$

   By taking the gradients of L with respect to W_1, W_2, X and setting them to zero, we get $I - 2Z_{11} = 0$, $I - 2Z_{22} = 0$, and $Z_{12} + Z_{21}^\top + \mathcal{A}^*(z) = 2Z_{12} + \mathcal{A}^*(z) = 0$. Therefore we have

$$\begin{aligned} \text{maximize} \quad & b^\top z \\ \text{subject to} \quad & \begin{bmatrix} \frac{1}{2}I & -\frac{1}{2}\mathcal{A}^*(z) \\ -\frac{1}{2}\mathcal{A}^*(z)^\top & \frac{1}{2}I \end{bmatrix} \succeq 0 \end{aligned}$$

   which is equivalent to[6] the condition stated on [RFP10, p. 480]. □

[6] Note that

$$A = \begin{bmatrix} I_m & Y \\ Y^\top & I_n \end{bmatrix} \succeq 0 \quad \text{is equivalent to} \quad B = \begin{bmatrix} I_m & -Y \\ -Y^\top & I_n \end{bmatrix} \succeq 0$$

6. (a) Consider some extreme examples of the *restricted isometry property* (RIP) [RFP10, Definition 3.1]. Assume $m = n = 2$ and let rank $r = 1$.

   i. Let $\mathcal{A}(X) = [x_{11}, x_{12}, x_{21}]'$. Find a matrix X which proves that there is no $\delta_r < 1$ that works, i.e., satisfies [RFP10, Eqn (3.2)], for all X.

   Solution. Let $X = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. We have

   $$\|X\|_F = \sqrt{x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2} = \sqrt{0^2 + 0^2 + 0^2 + 1^2} = 1$$

   and

   $$\|\mathcal{A}(X)\| = \|[0, 0, 0]'\| = \sqrt{0^2 + 0^2 + 0^2} = 0$$

   Plugging in the RIP inequality [RFP10, Eqn (3.2)], we get $\delta_r \geq 1$, implying that there is no $\delta_r < 1$ that works, as desired. □

   ii. Let $\mathcal{A}(X) = [x_{11}, x_{12} + x_{21}, x_{22}]'$. Complete the proof that $\delta_r = 1 - 1/\sqrt{2}$ works, and find an example X that shows that no smaller δ_r works.

   Proof. Given $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$, and since $r = 1$, X has the property $x_{11}x_{22} = x_{12}x_{21}$. Given

   $$\|X\|_F = \sqrt{x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2}$$

   and

   $$\|\mathcal{A}(X)\|_2 = \sqrt{x_{11}^2 + (x_{12} + x_{21})^2 + x_{22}^2}$$

   Note that $(x_{11} + x_{22})^2 + (x_{12} + x_{21})^2 \geq 0$ along with the fact that $x_{11}x_{22} = x_{12}x_{21}$ leads to $\|\mathcal{A}(X)\|_2^2 \geq \|X\|_F^2/2$. Also, $(x_{11} - x_{22})^2 + (x_{12} - x_{21})^2 \geq 0$ leads to $\|\mathcal{A}(X)\|_2^2 \leq 3\|X\|_F^2/2$.

   Given the bounds, we aim for δ_r such that:

   $$(1 - \delta_r)\|X\|_F \leq \|\mathcal{A}(X)\|_2 \leq (1 + \delta_r)\|X\|_F$$

   For the minimal δ_r that fits these inequalities, we need:

   $$(1 - \delta_r)^2\|X\|_F^2 \leq \|X\|_F^2/2 \quad \text{and} \quad (1 + \delta_r)^2\|X\|_F^2 \geq 3\|X\|_F^2/2$$

   Simplifying, we get $\delta_r \geq 1 - 1/\sqrt{2}$ as desired.

---

as the quadratic forms

$$q_A(x, y) = \begin{bmatrix} x^\top & y^\top \end{bmatrix} \begin{bmatrix} I_m & Y \\ Y^\top & I_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^\top x + x^\top Y y + y^\top Y^\top x + y^\top y = q_B(x, -y)$$

need to be non-negative for all $(x, y) \in \mathbb{R}^{m+n}$.

For example, consider $X = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$. We have $\|X\|_F = 2$ and $\|\mathcal{A}(X)\|_2 = \sqrt{2}$. Plugging in the RIP inequality [RFP10, Eqn (3.2)], we get $\delta_r \geq 1 - 1/\sqrt{2}$, implying that there is no smaller δ_r that works, as desired. □

(b) Complete the proof of [RFP10, Lemma 3.4].

Proof. Let $A, B \in \mathbb{R}^{m \times n}$ be matrices of the same dimensions. We want to show that there exist matrices B_1 and B_2 such that

i. $B = B_1 + B_2$,

ii. $\operatorname{rank} B_1 \leq 2 \operatorname{rank} A$,

iii. $AB_2^\top = 0$ and $A^\top B_2 = 0$,

iv. $\langle B_1, B_2 \rangle = 0$.

Consider a *full* singular value decomposition of A,

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^\top$$

where $\Sigma \in \mathbb{R}^{r \times r}$, and let $\hat{B} := U^\top B V$. Partition $\hat{B}$ as

$$\hat{B} = \begin{bmatrix} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & \hat{B}_{22} \end{bmatrix}$$

where $\hat{B}_{11} \in \mathbb{R}^{r \times r}$. Define now

$$B_1 := U \begin{bmatrix} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & 0 \end{bmatrix} V^\top, \quad B_2 := U \begin{bmatrix} 0 & 0 \\ 0 & \hat{B}_{22} \end{bmatrix} V^\top,$$

and we check that B_1 and B_2 satisfy the above conditions. Indeed, i, ii, iv are trivially true: $B = U\hat{B}V^\top = B_1 + B_2$, and since $U^\top U = V^\top V = I$ we have

$$AB_2^\top = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \hat{B}_{22}^\top \end{bmatrix} U^\top = U \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} U^\top = 0$$

and similarly $A^\top B_2 = 0$. Lastly, $\langle B_1, B_2 \rangle = \operatorname{tr}(B_1^\top B_2) = \operatorname{tr}(\mathbf{0}) = 0$.

We now show that $\operatorname{rank} B_1 \leq 2 \operatorname{rank} A = 2r$. Indeed, we have $\operatorname{rank}(\hat{B}_{11}) \leq r$ and $\operatorname{rank}(\hat{B}_{21}) \leq \min(m - r, r) \leq r$, $\operatorname{rank}(\hat{B}_{12}) \leq \min(r, n - r) \leq r$. Therefore,

$$\operatorname{rank}(B_1) \leq \min\{\operatorname{rank}(\hat{B}_{11}) + \operatorname{rank}(\hat{B}_{12}), \operatorname{rank}(\hat{B}_{11}) + \operatorname{rank}(\hat{B}_{21})\} \leq r + r = 2r,$$

as desired. □

- For $f(x) = a|x|, a \in \mathbb{R}$, consider Lipschitz constant $K > |a|$ and any neighborhood including 0.
- For $f(x) = \begin{cases} x^2\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$, note that previously we have proved that the derivative of f at $x = 0$ is 0. For any $x \neq 0$ we have $|f'(x)| \leq 2|x| \cdot \left|\sin\frac{1}{x}\right| + \left|\cos\frac{1}{x}\right| \leq 3$. Therefore we can pick any Lipschitz constant $K > 3$ and any neighborhood including 0.

For $f(x) =$ 3rd largest entry of $x, n \geq 3$, note that the perturbation is bounded by the largest change in any single element of x, which directly satisfies the Lipschitz condition $|f(x) - f(y)| \leq \|x - y\|$ for some $L > 1$.

For $f(x) =$ largest entry of Ax, where A is any $n \times n$ matrix, take any vectors x and y,

$$|f(x) - f(y)| = \left|\max_i(Ax)_i - \max_i(Ay)_i\right| \leq \max_i|(Ax)_i - (Ay)_i|$$
$$\leq \|Ax - Ay\| \leq \|A\|\|x - y\|$$

where $\|A\|$ is the operator norm of A. Hence, $f(x) = \max_i(Ax)_i$ is locally Lipschitz with a Lipschitz constant that could be the operator norm of A.

By [RW98, Thm 9.61], we have for regular at x functions $\partial^C f(x) = \partial f(x)$. Therefore,

- $f(x) = |x|^3$ has $\partial^C f(0) = \partial f(0) = \{0\}$,
- $f(x) = a|x|(a \geq 0)$ has $\partial^C f(0) = [-a, a]$,
- and $f(x) =$ largest entry of Ax has $\partial^C f(0) = \{A^\top y : y \in \text{conv}\{e_1, ..., e_n\}\}$ similarly.

On the other hand, by definition we have

- for $f(x) = a|x|(a < 0)$, $\partial^C f(0) = \text{conv}(\partial f(0)) = \text{conv}(\{-a, a\}) = [-a, a]$,
- for $f(x) = \begin{cases} x^2\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$, $\partial^C f(0) = \text{conv}([-1, 1]) = [-1, 1]$,
- for $f(x) =$ 3rd largest entry of x, $\partial^C f(0) = \text{conv}(\{y : y \in \text{conv}\{e^1, \ldots, e^n\}\}) = \text{conv}\{e^1, \ldots, e^n\}$.

$\square$

References

[BPC+11] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. doi:10.1561/2200000016.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi:10.1017/CBO9780511804441.

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, Nov 1995. doi:10.1145/227683.227684.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37:319–357, 2007. doi:10.1137/S0097539705447372.

[Lew99] A. S. Lewis. Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84(1):1–24, 1999. doi:10.1007/s10107980004a.

[Nes18] Yurii Nesterov. *Lectures on Convex Optimization*. Springer Cham, 2nd edition, 2018. doi:10.1007/978-3-319-91578-4.

[RFP10] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010. doi:10.1137/070697835.

[RW98] R. Tyrrell Rockafellar and Roger J. B. Wets. *Variational Analysis*. Springer Berlin, Heidelberg, 1998. doi:10.1007/978-3-642-02431-3.