On the Hardness of the $L_1 - L_2$ Regularization Problem

Kyle Yates

Clemson Graduate Student Seminar

August 29, 2024

Presentation Outline

- The Complexity Zoo
- How to get rich
- Signal Processing Compressed Sensing
- $L_1 - L_2$ Minimization

The Complexity Zoo

P. The class of P consists of problems that can be solved in polynomial time.

NP. A problem is called NP if there is a short witness when the answer is yes; that is, for the yes-instance of the problem, there is a solution which can be checked in polynomial time.

NP-complete. A problem in NP is called NP-complete if every problem in NP can be reduced to it in polynomial time.

NP-hard. A problem is called NP-hard if every problem in NP-complete can be reduced to it in polynomial time.

The Complexity Zoo (Simplified)

P. Problems that are easy to solve.

NP. Problems that are easy to verify a solution for.

NP-complete. The hardest problems in NP. Solving any NP-complete problem solves every problem in NP.

NP-hard. Problems that are at least as hard as NP-complete problems but may not be in NP. Solving any NP-hard problem solves every NP-complete problem.

How to Get Rich

Prove or disprove:

Problem (P vs. NP)

$P \neq NP$

This is one of the seven Millennium Prize Problems. Solving it comes with a $1 million dollar reward.

Examples

An NP-complete problem:

8					5			
	7		9				4	
		9		7	8	3	2	5
3		1		9			5	
		6				1		
	9			3		6		2
2	8	3	6	5		7		
	1				2		8	
			1					9

Solving an arbitrary $n^2 \times n^2$ sudoku grid of $n \times n$ blocks is NP-complete.

Examples

Another NP-complete problem:

It is NP-complete to assemble an optimal Bitcoin block.

Examples

An NP-hard problem:

It is NP-hard to decide whether the goal is reachable from the start of a stage in generalized Super Mario Bros [1].

Another Strategy to Get Rich

If you can design a fast algorithm for sudoku or Mario,
you can build optimal Bitcoin blocks and break most cryptosystems!

There is a deep fundamental connection between all these problems (and many more).

Partition Problem

Definition (Partition Problem)

Let $S = \{a_1, \ldots, a_n\}$ be a multiset of integers or rational numbers. Given S, find a partition S into two disjoint subsets S_1 and S_2 such that the sum of elements in S_1 is equal to the sum of elements in S_2.

Example

Consider $S = \{3, 1, 1, 2, 2, 1\}$. Let $S_1 = \{1, 1, 1, 2\}$ and $S_2 = \{2, 3\}$.

$$1 + 1 + 1 + 2 = 5$$
$$2 + 3 = 5$$

The elements of S_1 and the elements of S_2 both sum to 5 and form a partition of S.

The partition problem is NP-complete!

That's enough on complexity theory for now...
don't worry, it will be back!

Let's move on to our next topic:
optimization and signal processing.

L_0 Norm

Definition (L_0 Norm)

For a vector $x \in \mathbb{R}^n$, we define $\|x\|_0$ to be the number of nonzero entries in x. This is known as the L_0 norm (or Hamming weight if in $\mathbb{F}_q^n$).

Example

We have the following L_0 norms of vectors.

$$\|(1,0,0,1,1)\|_0 = 3$$
$$\|(\pi,0,0,e,0)\|_0 = 2$$
$$\|(0,0,\ldots,0)\|_0 = 0$$
$$\|(1,1,\ldots,1)\|_0 = n$$

Other Norms

Recall also the L_1, L_2, and L_p norms:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Compressed Sensing

An important problem in signal processing is known as the compressed sensing problem.

Problem (Compressed Sensing)

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, find the sparsest solution to the system $Ax = b$.

With $m < n$, compressed sensing is used to reconstruct sparse signals of length n from m samples. We can formulate this as a minimization problem using our L_0 norm:

$$\min_{x \in \mathbb{R}^n} \|x\|_0$$

$$\text{subject to } Ax = b$$

Unfortunately, compressed sensing is known to be NP-hard. To overcome this, various alternative problems have been proposed to approximate sparse solutions.

1. L_1 **Minimization:** $\min\{\|x\|_1 : Ax = b, x \in \mathbb{R}^n\}$
2. **Greedy Algorithms**
3. L_p **Minimization:** $\min\{\|x\|_p : Ax = b, x \in \mathbb{R}^n\}$ for $p < 1$
4. $L_1 - L_2$ **Minimization [2]:** $\min\{\|x\|_1 - \|x\|_2 : Ax = b, x \in \mathbb{R}^n\}$

We will be focusing on $L_1 - L_2$ minimization for the rest of this talk.

Is $L_1 - L_2$ Minimization Better than L_1?

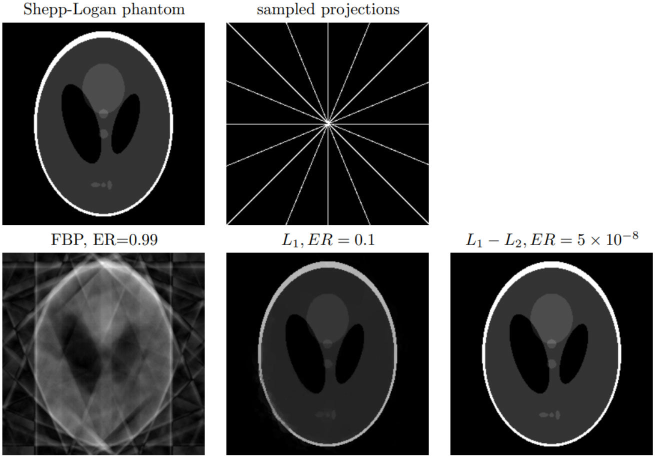

Fig. 6: MRI reconstruction results. It is demonstrated that 8 projections are enough to have exact recovery using $\ell_1 - \ell_2$. The relative errors are provided for each method.

Image from [3].

Recall the $L_1 - L_2$ minimization problem:

$$\min_{x \in \mathbb{R}^n} \|x\|_1 - \|x\|_2$$
$$\text{subject to } Ax = b$$

The constraints $Ax = b$ are nice! Unfortunately, the objective function $\|x\|_1 - \|x\|_2$ has some bad properties. It is:

- Nonlinear
- Non-convex
- Non-differentiable
- Non-separable

This makes our problem difficult to analyze and work with. Luckily, it has a few redeeming qualities. It is also:

- A difference of convex functions
- Lipschitz continuous
- $\|x\|_1 - \|x\|_2 \geq 0$

Main Result: NP-hardness of $L_1 - L_2$ Minimization

Theorem (Y. Ouyang, K. Y.)

The optimization problem

$$\min_{x \in \mathbb{R}^n} \|x\|_1 - \|x\|_2$$
$$\textit{subject to } Ax = b$$

is NP-hard.

This result is fairly involved to prove, so we'll prove it in a few pieces.

1. First, we'll prove an additional lemma.
2. Second, we'll prove an easier version of this problem with non-negative variables.
3. Finally, we'll extend the non-negative version to the general version.

Consider the minimization problem

$$\begin{aligned}\min_{x,y\in\mathbb{R}^n} f(x,y) = \sum_{i=1}^{n}(x_i+y_i) - \sqrt{\sum_{i=1}^{n} x_i^2 + y_i^2} \\ \text{subject to} \quad x_i + y_i = 1 \quad \text{for } i = 1,\dots,n \\ x_i, y_i \geq 0 \quad \text{for } i = 1,\dots,n\end{aligned} \tag{1}$$

Lemma (Y. Ouyang, K. Y.)

The set of optimal solutions to (1) is

$$X^* = \{(x,y)\in\mathbb{R}^{2n} \mid (x_i,y_i) = (0,1) \text{ or } (x_i,y_i) = (1,0)\ \forall i \in [n]\}$$

with optimal objective value $n - \sqrt{n}$.

Proof. Any feasible solution to (1) satisfies $x_i + y_i = 1$ for each i, meaning any feasible solution satisfies $\sum_{i=1}^{n}(x_i + y_i) = n$. So, (1) is equivalent to

$$\begin{aligned} \min_{x,y\in\mathbb{R}^n} h(x,y) = n - \sqrt{\sum_{i=1}^{n} x_i^2 + y_i^2} \\ \text{subject to} \quad x_i + y_i = 1 \quad & \text{for } i = 1,\ldots,n \\ x_i, y_i \geq 0 \quad & \text{for } i = 1,\ldots,n \end{aligned} \tag{2}$$

Since n is constant, we can find an optimal solution to (1) by solving

$$\begin{aligned} \max_{x,y\in\mathbb{R}^n} \quad & \sqrt{\sum_{i=1}^{n} x_i^2 + y_i^2} \\ \text{subject to} \quad & x_i + y_i = 1 \quad \text{for } i = 1, \ldots, n \\ & x_i, y_i \geq 0 \quad \text{for } i = 1, \ldots, n \end{aligned}$$

We have that

$$x_i^2 + y_i^2 \leq x_i + y_i = 1$$

with equality holding if and only if $(x_i, y_i) = (1, 0)$ or $(x_i, y_i) = (0, 1)$ for each i. Thus, the set of optimal solutions is X^*. The optimal objective value follows immediately. □

This will be enough to show that the nonnegative version of the problem is NP-hard.

Theorem (Y. Ouyang, K. Y.)

The optimization problem

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & \|x\|_1 - \|x\|_2 \\ \text{subject to } & Ax = b \\ & x \geq 0 \end{aligned} \tag{3}$$

is NP-hard.

We show this by providing a polynomial time reduction from the NP-complete partition problem to $L_1 - L_2$ minimization.

Proof. Suppose we have an instance of the partition problem with multiset $S = \{a_1, \ldots, a_n\}$ and let $a = (a_1, a_2, \ldots, a_n)$. Without loss of generality to dimension, let $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_n)$ be the vector of decision variables for (3) and define $A \in \mathbb{R}^{2n\times(n+1)}$ and $b \in \mathbb{R}^{(n+1)}$ as

$$A = \begin{bmatrix} I_n & I_n \\ a^T & -a^T \end{bmatrix}, \qquad b = \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix},$$

where I_n is an $n \times n$ identity matrix and $\mathbf{1}$ is a vector of n ones. Then, (3) is equivalent to (1) with additional constraint $a^T(x - y) = 0$. i.e., we have the problem

$$\begin{aligned} \min_{x,y\in\mathbb{R}^n} f(x, y) = \sum_{i=1}^{n}(x_i + y_i) - \sqrt{\sum_{i=1}^{n} x_i^2 + y_i^2} \\ \text{subject to} \quad x_i + y_i = 1 \quad \text{for } i = 1, \ldots, n \\ a^T(x - y) = 0 \\ x_i, y_i \geq 0 \quad \text{for } i = 1, \ldots, n \end{aligned} \tag{4}$$

If (3) has optimal solution (x, y) with objective value $f(x, y) = n - \sqrt{n}$, then we must have $a^T(x - y) = 0$ and $(x, y) \in X^*$. Let X be the set of indices such that $x_j = 1$ and $y_j = 0$, and let Y be the set of indices such that $x_j = 0$ and $y_j = 1$. Since $a^T(x - y) = 0$, we have

$$\sum_{j \in X} a_j = \sum_{j=1}^{n} a_j x_j = a^T x = a^T y = \sum_{j=1}^{n} a_j y_j = \sum_{j \in Y} a_j$$

which solves the partition problem. Likewise, if the partition problem has a solution, then, there is a vector $(x, y) \in X^*$ such that $a^T(x - y) = 0$.

On the other hand, suppose there is no solution to the partition problem. Then, there is no $(x, y) \in X^*$ such that $a^T(x - y) = 0$, and an optimal solution to (3) must have objective value strictly larger than $n - \sqrt{n}$ by our lemma. Likewise, if (3) has optimal solution (x, y) with objective value $f(x, y) > n - \sqrt{n}$, then $(x, y) \notin X^*$ and the partition problem has no solution. Thus, (3) is NP-hard. □

The last part of this proof is pretty dense. Think of it this way:

- **If a solution to the partition problem exists**, we can find it from solving this specific optimization problem.

- **If a solution to the partition problem does NOT exist**, we can also determine this from solving our optimization problem.

- If we can solve $L_1 - L_2$ minimization, we can solve the partition problem. This means $L_1 - L_2$ minimization is at least as hard as the partition problem.

Extension to the Non-negative Setting

We are off to a good start, but remember that we have not actually solved the original problem. We have shown NP-hardness when x is restricted to be non-negative – but we also want to show this result for the general case.

This is more difficult than we may anticipate. Recall that in a previous proof, we used the fact that

$$x_i^2 + y_i^2 \leq x_i + y_i = 1$$

But, this was only true since $0 \leq x_i, y_i \leq 1$. So, we can not use the same argument as before.

Consider the minimization problem

$$\min_{x,y\in\mathbb{R}^n} f(x,y) = \sum_{i=1}^{n} |x_i| + |y_i| - \sqrt{\sum_{i=1}^{n} x_i^2 + y_i^2} \tag{5}$$

$$\text{subject to} \quad x_i + y_i = 1 \quad \text{for } i = 1, \ldots, n$$

Lemma (Y. Ouyang, K. Y.)

The set of optimal solutions to (5) is

$$X^* = \{(x,y) \in \mathbb{R}^{2n} \mid (x_i, y_i) = (0,1) \text{ or } (x_i, y_i) = (1,0)\ \forall i \in [n]\}$$

with optimal objective value $n - \sqrt{n}$.

$\rightarrow$ **more OR background is needed for this result**

Proof. First, note that the case when $x \geq 0$ and $y \geq 0$ has already been discussed, which has optimal objective value $n - \sqrt{n}$. So, fix a feasible solution (x', y') to (5) such that $x'_j < 0$ for some fixed index j.

We claim that $f(x', y') > n - \sqrt{n}$. We'll show this by formulating an alternate optimization problem.

Let I be the set of indices such that $x'_i < 0$. Without loss of generality, we can assume all negative components are x'_i's, as if there is a negative component $y'_i < 0$, then simply swap x'_i and y'_i.

Note I is nonempty since $j \in I$. By the constraints of (5), it is clear we must have $y_i' > 1$ for all $i \in I$. For all $i \in I$, we write

$$\begin{aligned} x_i' &= -\epsilon_i', \\ y_i' &= 1 + \epsilon_i' \end{aligned}$$

for some $\epsilon_i' > 0$. For all $k \notin I$, we write

$$\begin{aligned} x_k' &= \epsilon_k', \\ y_k' &= 1 - \epsilon_k' \end{aligned}$$

for some $1 \geq \epsilon_k' \geq 0$ for all $k \notin I$. Let $\epsilon' = (\epsilon_1', \ldots, \epsilon_n') \in \mathbb{R}^n$ be the vector of these values.

Define the function $g : \mathbb{R}^n \to \mathbb{R}$ as

$$g(\epsilon) =$$

$$\sum_{i \in I}(\epsilon_i + 1 + \epsilon_i) + \sum_{i \notin I}(\epsilon_i + 1 - \epsilon_i) - \sqrt{\sum_{i \in I} \epsilon_i^2 + (1+\epsilon_i)^2 + \sum_{i \notin I} \epsilon_i^2 + (1-\epsilon_i)^2}$$

$$= n + 2\sum_{i \in I} \epsilon_i - \sqrt{\sum_{i \in I} \epsilon_i^2 + (1+\epsilon_i)^2 + \sum_{i \notin I} \epsilon_i^2 + (1-\epsilon_i)^2}$$

This may seem complicated, but this is just the function $f(x, y) = \sum_{i=1}^{n} |x_i| + |y_i| - \sqrt{\sum_{i=1}^{n} x_i^2 + y_i^2}$ replaced in terms of ϵ's.

Notice that $f(x', y') = g(\epsilon')$.

Remember: we wanted to show $f(x', y') > n - \sqrt{n}$.

Since $f(x', y') = g(\epsilon')$, it suffices now to show that $g(\epsilon') > n - \sqrt{n}$.

Now, we consider the problem

$$\begin{aligned} \min_{\epsilon\in\mathbb{R}^n} \ g(\epsilon) = n+2\sum_{i\in I}\epsilon_i - \sqrt{\sum_{i\in I}\epsilon_i^2 + (1+\epsilon_i)^2 + \sum_{i\notin I}\epsilon_i^2 + (1-\epsilon_i)^2} \\ \text{subject to } \epsilon_i \leq 1 \quad \text{for all } i \notin I \\ \epsilon_i \geq 0 \quad \text{for all } i = 1, \ldots, n \end{aligned} \tag{6}$$

Note that ϵ' is feasible to (6). Any optimal solution to (6) will have $\epsilon_i \in \{0, 1\}$ for $i \notin I$, as we can maximize the term

$$\sum_{i\notin I}\epsilon_i^2 + (1-\epsilon_i)^2$$

with $0 \leq \epsilon_i \leq 1$ for $i \notin I$ independent from the rest of the problem.

Let $I^c = [n] \setminus I$. Therefore, we can solve (6) by instead looking at the problem

$$\min_{\epsilon \in \mathbb{R}^{|I|}} \; h(\epsilon) = n + 2\sum_{i \in I} \epsilon_i - \sqrt{\sum_{i \in I} \epsilon_i^2 + (1 + \epsilon_i)^2 + |I^c|} \quad \text{subject to } \epsilon_i \geq 0 \quad \text{for } i \in I \tag{7}$$

We will now show that $g(\epsilon') > n - \sqrt{n}$ by showing that the optimal objective value to (6) and (7) is $n - \sqrt{n}$, and also that ϵ' is not an optimal solution to (6). For the constraints $\epsilon_i \geq 0$ for $i \in I$ of (7), let $\{\mu_k\}_{k \in I}$ be the associated Lagrange multipliers. The KKT FONC are

$$2 - \frac{1 + 2\epsilon_k}{\sqrt{\sum_{i \in I} \epsilon_i^2 + (1 + \epsilon_i)^2 + |I^c|}} - \mu_k = 0 \text{ for } k \in I$$

$$\epsilon_k(-\mu_k) = 0 \text{ for } k \in I$$

We now consider cases. First, suppose that there is a KKT point ϵ with $\epsilon_k = 0$ for all $k \in I$. Then, we have

$$\sqrt{\sum_{i \in I} \epsilon_i^2 + (1 + \epsilon_i)^2 + |I^c|} = \sqrt{|I| + |I^c|} = \sqrt{n}$$

The gradient condition then becomes

$$2 - \frac{1}{\sqrt{n}} - \mu_k = 0 \text{ for } k \in I$$

Solving for μ_k, this gives

$$\mu_k = 2 - \frac{1}{\sqrt{n}} \text{ for } k \in I$$

Note that since $n \geq 1$, $\mu_k > 0$ for all $k \in I$, which is dual feasible. So, there always exists a KKT point of this form, which has objective value $n - \sqrt{n}$.

Now, suppose there is a KKT point ϵ with some $k \in I$ such that $\epsilon_k > 0$. Then, $\mu_k = 0$. The gradient condition for this k is

$$2 - \frac{1 + 2\epsilon_k}{\sqrt{\sum_{i \in I} \epsilon_i^2 + (1 + \epsilon_i)^2 + |I^c|}} = 0$$

$$\Rightarrow \sqrt{\sum_{i \in I} \epsilon_i^2 + (1 + \epsilon_i)^2 + |I^c|} = \frac{1 + 2\epsilon_k}{2}$$

In this case, the objective value at the point ϵ must be

$$\begin{aligned} g(\epsilon) &= n + 2\sum_{i \in I} \epsilon_i - \sqrt{\sum_{i \in I} \epsilon_i^2 + (1+\epsilon_i)^2 + |I^c|} \\ &= n + 2\sum_{i \in I} \epsilon_i - \frac{1+2\epsilon_k}{2} \\ &= n + \epsilon_k + 2\sum_{\substack{i \in I \\ i \neq k}} \epsilon_i - \frac{1}{2} \\ &> n - \sqrt{n} \end{aligned}$$

The last inequality follows from the fact that $\epsilon_i \geq 0$ for all $i \in I$ and $n \geq 1$.

The last inequality follows from the fact that $\epsilon_i \geq 0$ for all $i \in I$ and $n \geq 1$. Thus, the set of optimal solutions to (6) is given by

$$\mathcal{E}^* = \left\{ \epsilon \in \mathbb{R}^n_+ : \begin{cases} \epsilon_k & = 0 \text{ if } k \in I \\ \epsilon_k & \in \{0, 1\} \text{ if } k \notin I \end{cases} \right\}$$

Now, observe that $\epsilon' \notin \mathcal{E}^*$ since $j \in I$ and $\epsilon'_j > 0$, so $f(x', y') = g(\epsilon') > n - \sqrt{n}$. Hence, (x', y') is not optimal to (5). Since (x', y') was any arbitrary feasible solution to (5) with at least one negative component, no feasible solutions of this type are optimal. Thus, the set of optimal solutions to (5) is

$$X^* = \{(x, y) \in \mathbb{R}^{2n} \mid (x_i, y_i) = (0, 1) \text{ or } (x_i, y_i) = (1, 0) \; \forall i \in [n]\}$$

□

Final Result

Proving this optimal set is the difficult part. Once that we have it, our main theorem follows just as before.

Theorem (Y. Ouyang, K. Y.)

The optimization problem

$$\min_{x\in\mathbb{R}^n} \|x\|_1 - \|x\|_2$$
$$\textit{subject to } Ax = b$$

is NP-hard.

Other Results

In this work, we also consider the unconstrained $L_1 - L_2$ problem:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda(\|x\|_1 - \|x\|_2)$$

for penalty parameter $\lambda > 0$. This problem turns out to also be NP-hard for some λ (which is much more difficult to prove).

Thank you!

References

Aloupis G., Demaine E. D., Guo A., Viglietta G. Classic Nintendo games are (computationally) hard. Theoretical Computer Science, Volume 586, 2015, Pages 135-160, ISSN 0304-3975, https://doi.org/10.1016/j.tcs.2015.02.037.

Esser E., Lou Y., Xin J. A Method for Finding Structured Sparse Solutions to Nonnegative Least Squares Problems with Applications. SIAM Journal on Imaging Sciences, Volume 6, 2013, Pages 2010-2046.

Yin P., Lou Y., He Q., Xin J. Minimization of ℓ_{1-2} for Compressed Sensing. SIAM Journal on Scientific Computing, Volume 37, 2015, Pages A536-A563.

Gao, Shuhong. Lecture notes, Math 9850 Fall 2023 (Post-Quantum Cryptography).