باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

بهینه‌سازی در علوم داده

موضوع: NRP & Svdittik

نگارش:

امیرحسین جوادی

۹۷۱۰۱۴۸۹

علیرضا طبیب‌پور

۴۰۰۰۱۱۱۲۲

استاد راهنما:

دکتر مجتبی تفاق

تیر ۱۴۰۱

This page would be intentionally left blank if we would not wish to inform about that.

فهرست مطالب

This page would be intentionally left blank if we would not wish to inform about that.

فصل ۱

The Nurse Rostering Problem

۱.۱ مقدمه

ارکان سلامت و بیمارستان، به خصوص پرستاران، نقش بسیار مهمی در کیفیت زندگی انسان‌ها دارند و برنامه‌ریزی مناسب برای تعیین زمان کاری آنان به صورتی که هم فشار و استرس کاری زیادی بر آنان وارد نشود و هم سرویس مورد نیاز به بیماران داده شود، بسیار مهم است.

بعد از همه گیری کرونا، تقاضا روز افزونی از پرستاران برای کار به صورت اضافه کار به وجود آمد تا بیمارستان بتوانند به حجم زیاد بیماران کرونایی سرویس دهند. در این سناریو، روش‌های استاندارد زمان‌بندی جوابگو نخواهند بود چون امکان اضافه‌کاری در آن‌ها در نظر نگرفته شده‌است. به این منظور، فورمول بندی مسئله‌ی زمان‌بندی کاری پرستاران با امکان اضافه کاری به مراکز درمانی این قابلیت را می‌دهد تا با کمک این سیستم هوشمند خودکار، شیفت‌های پوشش داده نشده را با حداقل فشار کاری به پرستاران خدمت ارائه دهند.

پرستاران عدالت کاری و حجم زمانی متعادل را برای رفاه خود ضروری می‌دانند. برنامه‌ریزی مناسب پرسنل پرستاری می‌تواند با کاهش فشار کاری نامتعادل یا استرس مفرط که ممکن است منجر به کاهش کارایی پرستاران و خطاهای انسانی احتمالی شود، تأثیر زیادی بر کیفیت خدمات بهداشتی ارائه شده داشته باشد.

مسئله‌ی (The Nurse Rostering Problem (NRP حل مسئله‌ی انتساب پرستاران به شیفت کاری است تا برخی از محدودیت ها مانند تضمین حداقل سطح کمک مورد نیاز به بیماران را در عین بهینه سازی یک تابع هدف مانند تعداد کلی ساعات کار هر پرستار برآورده کند. این یک مسئله‌ی بهینه‌سازی NP-complete است. روش‌های گذشته‌ی حل این مسئله الگوریتم‌هایی مانند الگوریتم ژنتیک، Tabu Search و الگوریتم‌های Ant Colony Bases هستند.

در این پژوهش، ما یک فرمول برنامه‌ریزی انعطاف پذیری را تعریف می‌کنیم که قادر به مقابله با سناریوهای همه‌گیری است که در آن تعداد پرستاران ممکن است برای تضمین حداقل سطح کمک مورد نیاز کافی نباشد. هدف اصلی این پژوهش ارائه تسهیلات بهداشتی و درمانی با یک ابزار برنامه ریزی انعطاف پذیر و آسان است که با در نظر گرفتن امکان اضافه کاری پرستاران، متعادل ترین حجم کاری مورد نیاز پرستاران را ارائه کند.

۲.۱ NRP basic

همان طور که توضیح داده شد، هدف از این پروژه، طراحی سیستمی برای برنامه ریزی شیفت‌های کاری پرستان است؛ به گونه‌ای که در مواقع نیاز بسیار زیاد و بیش از ظرفیت نیز بتوان به بیماران رسیدگی کرد و همچنین فشار و استرس کاری پرستاران را کمینه کرد. برای طراحی چنین سیستمی، اول حالت شرایط عادی را بررسی می‌کنیم که تعداد پرستاران برای رسیدگی به بیماران کافی است. مسئله به این صورت است که N پرستار داریم و که برای یک دوره T روزه، قصد برنامه ریزی برای تعیین شیفت‌های کاری این پرستاران را داریم. هر روز ۳ شیفت صبح و عصر و شب دارد و شیفت s مقدار h_s ساعت است و تعداد R_s پرستار نیاز دارد.
برای تعریف ریاضی مسئله متغیرهای باینری x_{ist} و p_i را به صورت زیر تعریف می‌کنیم.

$$x_{ist} = \begin{cases} 1 & \text{if nurse } i\text{th covers shift } s\text{th on day } t\text{th} \\ 0 & \text{otherwise} \end{cases}$$

$$p_i = \begin{cases} 1 & \text{if the } i\text{th nurse worked on the last shift of the previous period} \\ 0 & \text{otherwise} \end{cases}$$

هدف مقاله اولیه، کمینه کردن مجموعساعت‌های کاری پرستاران است. پس مسئله بهین سازی به صورت زیر در می‌آید:

$$\min_{x_{ist}\in\{0,1\}} \sum_{i=1}^{N}\sum_{s=1}^{3}\sum_{t=1}^{T} x_{ist}h_s \tag{1.1}$$

$$\text{subject to} \sum_{s=1}^{3} x_{ist} \le 1 \qquad (\forall i = 1,\dots,N\ t = 1,\dots,T) \tag{2.1}$$

$$\sum_{i=1}^{N} x_{ist} = R_s \qquad (\forall s = 1,\dots,3\ t = 1,\dots,T) \tag{3.1}$$

$$\sum_{s=1}^{3}\sum_{t=1}^{T} x_{ist}h_s \ge H_i \qquad (\forall i = 1,\dots,N) \tag{4.1}$$

$$\sum_{s=1}^{3}\sum_{t=1}^{T} x_{ist}h_s \le H^{max} \qquad (\forall i = 1,\dots,N) \tag{5.1}$$

$$x_{i3t} + \sum_{s=1}^{3} x_{ist+1} \le 1 \qquad (\forall i = 1,\dots,N\ t = 1,\dots,T-1) \tag{6.1}$$

$$\sum_{s=1}^{3} x_{is1} \le (1 - p_i) \qquad (\forall i = 1,\dots,N) \tag{7.1}$$

که عبارت ۱.۱ همان مجموع ساعت‌های کاری پرستاران است. حال قیود را بررسی می‌کنیم:

- قید ۲.۱ این مفهوم را دارد که هر پرستار در هر روز حداکثر می‌توان یک شیفت کار کند.
- قید ۳.۱ این مفهوم را دارد که در شیفت s در هر روز، تعداد R_s پرستار نیاز است.
- قید ۴.۱ این مفهوم را دارد که شیفت s در هر روز، h_s ساعت است.

- قید ۵.۱ این مفهوم را دارد که هر پرستار در مجموع نباید بیشتر از H_{max} ساعت کار کند.
- قید ۶.۱ این مفهوم را دارد که اگر پرستاری در شیفت شب کار کند، نباید در هیچ شیفتی از روز بعدی کار کند.
- قید ۷.۱ همان مفهوم قبلی را دارد با این تفاوت که پرستارانی که در شیفت شب آخرین روز دوره قبل کار کرده‌اند، جرو داده‌های مسئله است و برحسب p_i به دست می‌آید.

پس یک مسئله linear mixed integer است. اما ما مسئله را کلی‌تر حل می‌کنیم و به چای کمینه کردن مجموع ساعت‌ها، مجموع دستمزدها را کمینه می‌کنیم. برای این منظور بردار w را به عنوان ورودی نیاز داریم که بردار دستمزد ساعتی پرستاران است و اگر بخواهیم همان مسئله قبلی را حل کنیم، کافی است بردار با تمام درایه‌های ۱ را به جای w قرار دهیم. پس مسئله تبدیل می‌شود به:

$$
\begin{aligned}
&\min_{x_{ist}\in\{۰,۱\}} \sum_{i=۱}^{N}\sum_{s=۱}^{۳}\sum_{t=۱}^{T} x_{ist}h_s w_i \\
\text{subject to } &\sum_{s=۱}^{۳} x_{ist} \leq ۱ \qquad (\forall i = ۱,\dots,N\ t = ۱,\dots,T) \\
&\sum_{i=۱}^{N} x_{ist} = R_s \qquad (\forall s = ۱,\dots,۳\ t = ۱,\dots,T) \\
&\sum_{s=۱}^{۳}\sum_{t=۱}^{T} x_{ist}h_s \geq H_i \qquad (\forall i = ۱,\dots,N) \\
&\sum_{s=۱}^{۳}\sum_{t=۱}^{T} x_{ist}h_s \leq H^{max} \qquad (\forall i = ۱,\dots,N) \\
&x_{i۳t} + \sum_{s=۱}^{۳} x_{ist+۱} \leq ۱ \qquad (\forall i = ۱,\dots,N\ t = ۱,\dots,T-۱) \\
&\sum_{s=۱}^{۳} x_{is۱} \leq (۱ - p_i) \qquad (\forall i = ۱,\dots,N)
\end{aligned}
$$

اما اگر در شرایط اضطراری تعداد پرستاران برای برآورده کردن قیود بالا کافی نباشد، مسئله‌های بالا infeasible می‌شوند. در این حالت مسئله را در بخش بعدی با معرفی مسئله در شرایط اضطراری حل می‌کنیم.

۳.۱ NRP: Formulation in Emergency Scenarios

در این حالت می‌خواهیم مسئله را از حالت $infeasble$ خارج کنیم. برای این کار، متغیر α_{st} را معرفی می‌کنیم که عددی نامنفی است و تعداد پرستارانی است که برای شیفت s در روز t کم داریم. (با این که ظاهرا عددی حسابی است اما می‌توانیم به صورت پیوسته تعریفش کنیم و در قیودی که برای آن می‌نویسیم، خودش عددی حسابی خواهد شد و در محاسبتمان صرفه جویی کنیم)

حال برای این که مسئله را $feasiible$ کنیم، قید ۳.۱ را به صورت زیر تبدیل میکنیم:

$$\sum_{i=۱}^{N} x_{ist} + \alpha_{st} = R_s \qquad (\forall s = ۱, \dots, ۳ \ t = ۱, \dots, T)$$

و همچنین در objective ترم زیر را اضافه می‌کنیم.

$$\rho_۱ \sum_{s=۱}^{۳} \sum_{t=۱}^{T} \alpha_{st}$$

دقیت کنید زمانی که می‌توان بدون کمبود پرستاران مسئله را حل کرد باید α صفر باشد. پس مقدار $\rho_۱$ باید بیشتر از ماکسیمم مقدار کل حقوق ممکن در یک شیفت باشد. (برای مسئله‌ای که خود مقاله حل کرده است کافی است که از ماکسیمم تعداد ساعت‌های شیفت‌ها بیشتر باشد)

این حالت مسئله feasible می‌شود ولی کمک چندانی نمی‌کند زیرا نمی‌توان سرویس مناسب را به بیماران داد. برای حل مشکل، در بخش بعد با معرفی شیفت اضافه کاری پرستاران، مشکل را برطرف می‌کنیم.

۴.۱ معرفی شیفت اضافه کاری

در شرایط اضطراری مانند همه گیری کرونا، به پرستاران اجازه کار در بیش از یک شیفت در یک روز داده می‌شود. اما باید بسیار محتاط برای کم شدن فشار کاری به رستاران باشیم. برای این مقصود، مفاهیم زیر را تعریف می‌کنیم.

پرستاران می‌توانند به اندازه کسر c از شیفت بعد از شیفت عادی خود را کار کنند و همجنین می‌توانند به اندازه کسر $۱ - c$ از شیفت قبل از شیفت عادی خود را کار کنند.

حال متغیرهای باینری q_{ist} و z_{ist} را تعریف می‌کنیم:

$$q_{ist} = \begin{cases} ۱ & \text{if nurse } i\text{th works the last } (۱-c)hs \text{ hours of the shift } s\text{th on day } t\text{th} \\ ۰ & \text{otherwise} \end{cases}$$

$$z_{ist} = \begin{cases} ۱ & \text{if nurse } i\text{th works the first } chs \text{ hours of the shift } s\text{th on day } t\text{th} \\ ۰ & \text{otherwise} \end{cases}$$

و مسئله به صورت زیر تبدیل می‌شود:

$$\min_{x_{ist}\in\{0,1\}} \sum_{i=1}^{N}\sum_{s=1}^{3}\sum_{t=1}^{T} h_s w_i(x_{ist}+cz_{ist}+(1-c)q_{ist}) + \rho_1 \sum_{s=1}^{3}\sum_{t=1}^{T}\alpha_{st} \tag{8.1}$$

$$\text{subject to} \sum_{s=1}^{3} x_{ist} \leq 1 \qquad (\forall i=1,\dots,N\ t=1,\dots,T) \tag{9.1}$$

$$\sum_{i=1}^{N}(x_{ist}+z_{ist})+\alpha_{st}=R_s \qquad (\forall s=1,\dots,3\ t=1,\dots,T) \tag{10.1}$$

$$\sum_{s=1}^{3}\sum_{t=1}^{T} x_{ist}h_s \geq H_i \tag{11.1}$$

$$x_{i3t}+\sum_{s=1}^{3} x_{ist+1} \leq 1 \tag{12.1}$$

$$\sum_{s=1}^{3} x_{is1} \leq (1-p_i) \tag{13.1}$$

$$z_{ist} \leq x_{is-1t} \tag{14.1}$$

$$q_{ist} \leq x_{is+1t} \tag{15.1}$$

$$z_{ist} \leq 1-x_{ist} \tag{16.1}$$

$$q_{ist} \leq 1-x_{ist} \tag{17.1}$$

$$\sum_{s=1}^{3}\sum_{t=1}^{T} h_s(x_{ist}+cz_{ist}+(1-c)q_{ist}) \leq H^{max} \tag{18.1}$$

$$\sum_{s=1}^{3}(z_{ist}+q_{ist}) \leq 1 \tag{19.1}$$

$$\sum_{i=1}^{N} z_{ist} = \sum_{i=1}^{N} q_{ist} \tag{20.1}$$

$$z_{i3t}+x_{i1t+1} \leq 1 \tag{21.1}$$

$$q_{i3t}+x_{i2t} \leq 1 \tag{22.1}$$

$$x_{ist}, z_{ist}, q_{ist} \in \{0,1\}, \alpha_{st} \in R \tag{23.1}$$

قیدهای ۹.۱ تا ۱۳.۱ و قید ۱۸.۱ را در بخشهای قبلی توضیح داده‌ایم. بقیه قیدها عبارت اند از:

- قید ۱۴.۱ و ۱۵.۱ به معناست که ساعت‌های اضافه کاری با ساعت‌های کاری شیفت عادی باید ترکیب شوند. همچنین داریم

$$x_{i4t}=x_{i1t+1} \qquad x_{i0t}=x_{i3t-1}$$

و همچنین فرض می‌کنیم که در اولین شیفت از دوره‌ی بعدی، هیچ پرستاری کار نمی‌کند. همچنین دقت کنید که $x_{i01}=p_i$ است.

- قید ۱۶.۱ و ۱۷.۱ به این معناست که در خود شیفت‌های کاری عادی، نمی‌توانیم اضافه کاری داشته باشیم.

- قید ۱۹.۱ به این معناست که هر پرستار در هر روز فقط یک بار می‌تواند اضافه کاری کند.
- قید ۲۰.۱ به این معناست که در هر شیفت تعداد پرستارانی که اضافه کاری قبل از شیفا عادیشان و بعد از شیفت عادیشان را می‌کنند برابر است. (که شیفت کسری نداشته باشیم و تمام شیفتاها به صورت کامل پر شوند)
- قید ۲۱.۱ به این معناست که اگر پرستاری در شیفت شب اضافه کاری بعد از شیفت عدای خودش را داشته باشد، صبح فردا نباید کار کند.
- قید ۲۲.۱ به این معناست که اگر پرستاری در شیفت شب اضافه کاری قبل از شیفت عادی خود کار کند، شیفت عصر آن روز نباید کار کند.

نکته: این مسئله کاملا قابل گسترش است و می‌توان آن را به حالت‌های کلی تری که بیمارستان علاقه دارد و قیدهای دیگری را اضافه یا کاهش دهد، تبدیل کرد.

۵.۱ نتایج عددی

در این قسمت، نتایج عددی دو الگوریتم توضیح داده شده را با یک دیگر مقایسه می‌کنیم. این مقایسه را هم در حالت مساوی بودن حقوق پرستاران و هم در حالت حقوق متفاوت بررسی خواهیم کرد. در الگوریتم اول اضافه‌کاری امکان نداشت ولی در الگوریتم دوم این امکان به الگوریتم اضافه شده است. الگوریتم اول با نام Base در فایل NRP_Base.ipynb و الگوریتم دوم با نام Ext در فایل NRP_Ext.ipynb در پوشه‌ی کدها وجود دارد.
این مقایسه را در حالت‌های متفاوت تعداد پرستاران و تعداد روزها انجام خواهیم داد. در این آزمایش عددی، موارد زیر فرض شده‌اند.

$$k = [\frac{T}{۷}]$$

$$H_i = ۳۶k \quad H_{\max} = ۶۰ \quad c = ۰/۵$$

$$x_{i۴T} = ۰ \quad x_{i۰۱} = ۰$$

$$R = [۶, ۵, ۴] \quad h = [۷, ۸, ۹] \quad p = [۱, ۰, ۰, \dots, ۰]$$

این آزمایش در بیشترین حالت به تعداد $۳ \times N \times T \times ۳$ متغیر باینری و به تعداد $۳ \times T$ متغیر حقیقی دارد.
در این آزمایش عددی از زبان برنامه‌نویسی Julia و پکیج JuMP و GLPK Solver که توانایی حل مسائل linear mixed integer را دارد، استفاده شده‌است. در جدول زیر در حالتی که حقوق پرستاران مساوی فرض شده‌است، (یعنی مانند مقاله به دنباله کمینه کردن ساعت‌های کاری پرستاران هستیم) موارد زیر را برای کدهای Base و Ext بررسی کرده‌ایم.

۱. زمان مصرف شده در CPU برای نتیجه دادن Solver
۲. بررسی این که آیا جواب به دست آمده بهینه است یا نه.

۳. تعداد مواردی که پرستاران برای پوشش کامل شیفت‌ها کافی نیستند

۴. بیشترین تعداد کمبود پرستاران در بین همه‌ی شیفت‌ها

۵. مجموع تعداد کمبود پرستاران در بین همه‌ی شیفت‌ها

۶. بیشترین تعداد ساعت کاری بین پرستاران در کل بازه‌ی بررسی شده

N	T	code	CPU Time	Optimal	Not enough cases	maximum of lack of nurses	Sum of lack nurses	Maximum total hours worked
15	7	Base	0.017	1	5	4	13	55
15	7	Ext	3.385	1	0	0	0	60
15	14	Base	0.019	1	14	3	28	106
15	14	Ext	15.58	1	0	0	0	119

جدول ۱.۱: حل مسئله برای کمترین ساعت کاری

همین مقایسه را در حالتی که حقوق پرستاران با یک دیگر مساوی نیست در جدول زیر نشان داده شده است. (یعنی به دنبال کمینه کردن هزینه‌ی حقوق پرستاران هستیم)

N	T	code	CPU Time	Optimal	Not enough cases	maximum of lack of nurses	Sum of lack nurses	Maximum total hours worked
15	7	Base	0.034	1	6	4	13	55
15	7	Ext	7.72	1	0	0	0	60
15	14	Base	0.052	1	14	3	28	106
15	14	Ext	42.16	1	0	0	0	119

جدول ۲.۱: حل مسئله برای حل کمترین هزینه‌ی حقوق

از جدول‌های بالا به این نتیجه می‌رسیم که کد Base به صورت قابل انتظاری زمان کمتری برای حل مسئله نیاز دارد. اما کد Ext برای پوشش همه‌ی شیفت‌ها بسیار موثرتر است و مقادیر α را به صفر می‌رساند و در نتیجه، کارایی خودش را برای مواقع اضطراری که با کمبود تعداد پرستاران روبرو هستیم نشان می‌دهد. در حالت متفاوت بودن حقوق پرستاران، رابطه‌ی بین کد Base و Ext به صورت مشابه است و همچنین زمان مورد نیاز برای حل مسئله به طرز قابل ملاحظه‌ای بیشتر می‌شود.
در ادامه، کدهای Base و Ext در یک حالت $T = ۷$ و $N = ۱۵$ با جزئیات بیشتر مقایسه می‌کنیم.

Nurse	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Total shifts	6	5	4	6	4	6	7	7	6	7	7	7	6	7	7
Total hours	44	40	36	44	36	46	51	51	47	54	54	54	48	52	52

جدول ۳.۱: جزئیات حل مسئله برای حل کمترین ساعت کاری

Nurse	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Total shifts	5	5	5	7	5	6	6	5	5	5	5	5	5	5	5
Extra shifts after regular shifts	2	1	1	1	2	1	2	0	3	2	4	3	1	2	1
Extra shifts before regular shifts	3	2	3	0	2	2	1	1	1	2	1	2	3	1	2
Total hours in regular shifts	37	42	42	49	39	47	43	39	38	37	39	40	41	41	39
Extra hours	20.5	12.5	15	8.5	8.5	11.5	17	3.5	17.5	17	20	20	16	13	12

جدول ۴.۱: جزئیات حل مسئله برای حل کمترین هزینه‌ی حقوق

۶.۱ نتیجه گیری

در این پژوهش، سعی کردیم مسئله‌ی Nurse Rostering Problem را با استفاده از بهینه‌سازی linear mixed integer حل کنیم. مسئله را در دو حالت امکان اضافه کاری و عدم امکان اضافه کاری با یک دیگر بررسی کردیم و همچنین مسئله را در حالت کلی‌تری نسبت به مقاله مورد بررسی قرار دادیم به نحوی که بتوانیم مجموع ساعت کاری پرستاران یا هزینه‌ی حقوق پرستاران را کمینه کنیم. نتایج عددی را مورد بررسی قرار دادیم و کارایی دو الگوریتم را با هم مقایسه کردیم.

فصل ۲

Iterated Tikhonov Regularization

۱.۲ مقدمه

معادلات خطی و مسائل کمترین مربعات نامساعد، ill-conditioned در زمینه‌های مهندسی و علمی بسیار پرکاربرد است. این روش‌ها می‌توانند به خطای داده به طور مثال به خاطر data perturbation، بسیار حساس باشد. به همین دلیل روش های regularization سعی می‌کند که حساسیت این مسائل را با جایگینی به یک مسئله نزدیک دیگر کمتر کند.

این پژوهش توضیح می‌دهد که چگونه تجزیه مقدار ویژه‌ی تعمیم‌یافته را می‌توان با ورژن تکرار شونده‌ی روش Tikhonov regularization ترکیب کرد و نشان می‌دهد که روش به‌دست‌آمده، راه‌حل‌های تقریبی با کیفیت بالاتر را نسبت به روش رایج‌تر استفاده‌شده برای تجزیه مقدار ویژه‌ی تعمیم‌یافته با Tikhonov regularization استاندارد تعیین می‌کند.

۲.۲ تعریف مسئله

مسائل کمترین مربعات خطی را که به صورت زیر است در نظر بگیرید.

$$\min_{x \in R^n} ||Ax - b||; \tag{۱.۲}$$

که ماتریس $A^{m\times n}$ مقدار ویژه‌های کوچک دارد. فرض بر این است که بردار داده $b \in R^m$ توسط یک خطای ناشناخته $e \in R^m$ آلوده شده است که ممکن است ناشی از عدم دقت اندازه گیری یا گسسته سازی باشد. مسایل حداقل مربعات با ماتریسی از این نوع به عنوان مسائل گسسته خطی ill-conditioned نامیده می شوند.
اجازه دهید b_{true} بردار مجهول بدون خطا را نشان دهد به طوری که

$$b = b_{true} + e$$

فرض می‌کنیم که بردار b_{true} در برد A است و یک کران δ برای نرم خطا می‌دانیم.

$$||e|| \leq \delta$$

ما می خواهیم تقریب دقیقی از پاسخ x_{true} در مسئله‌ی ۱.۲ را با داشتن b به جای b_{true} به دست بیاوریم. می‌توانیم x_{true} را به صورت زیر بیان کنیم.

$$x_{true} = A^{\dagger} b_{true}$$

که $A^{\dagger}$ pseudoinverse ماتریس A است. با توجه به خطای e در b و وجود مقدار ویژه‌های نزدیک به مبدا ماتریس A، جواب $A^{\dagger}b$ برای مسئله‌ی ۱.۲ عموماً تقریب معنی‌داری از x_{true} نیست. برای حل این مشکل، به طور معمول مشکل کمینه سازی ۱.۲ را با یک مسئله مرتبط جایگزین می‌کنیم که راه حل آن نسبت به خطای b حساسیت کمتری دارد. این جایگزینی با عنوان regularization شناخته می‌شود.

$$\min_{x \in R^n} \{||Ax - b||^۲ + \mu ||L(x - x^{(۰)})||^۲\} \tag{۲.۲}$$

که $x^{(۰)}$ ممکن است یک تقریب در دسترس از x_{true} باشد. اگر هیچ تقریبی از x_{true} مشخص نباشد، می توان $x^{(۰)} = ۰$ را انتخاب کرد. ماتریس $L \in R^{p\times n}$ به عنوان یک ماتریس regularization شناخته می‌شود. انتخاب های رایج L شامل ماتریس همانی I یا تقریب گسسته از یک عملگر دیفرانسیل است. ما نیاز داریم که L شرط زیر را ارضا کند.

$$N(A) \cap N(L) = ۰$$

که $N(M)$ برابر null space ماتریس M است. پس حل مسئله‌ی ۲.۲ به شکل زیر برای هر $\mu > ۰$ است.

$$x_\mu = (A^T A + \mu L^T L)^{-۱}(A^T b + L^T L x^{(۰)})$$

پارامتر رگولاریزیشن $\mu > ۰$ حساسیت x_μ را به خطای e در b را تعیین می‌کند و نزدیکی x_μ به بردار مطلوب x_{true} تاثیرگذار است. روش‌های زیادی برای مشخص کردن پارامتر رگولاریزیشن مناسب وجود دارد مانند discrepancy principle، L-curve criterion و generalized cross validation که نیازمند تکرار حل مسئله‌ی ۲.۲ برای تعداد

زیادی μ است. برای این مسئله از تجزیه‌ی مقدار ویژه‌ی تعمیم یافته‌ی جفت ماتریس $\{A, L\}$ استفاده می‌کنیم. اعمال این تجزیه روی مسئله‌ی ۲.۲ مسئله کمینه سازی را به یک مسئله با ماتریس های قطری تبدیل می‌کند، که حل مکرر مسئله تبدیل شده را برای دنباله‌ای از مقادیر μ از لحاظ محاسباتی مورد قبول می‌کند.
این پژوهش پیشنهاد می‌کند که iterated Tikhonov regularization با Tikhonov regularization جایگزین شود. تلاش محاسباتی اضافی مورد نیاز در مقایسه با محاسبات مورد نیاز برای محاسبه GSVD جفت ماتریس $\{A, L\}$ ناچیز است.

۳.۲ Iterated Tikhonov regularization based on the GSVD

فرض کنید $x^{(۰)} = x_\mu$ یک تقریب در دسترس از x_{true} باشد، مثلا پاسخ مسئله‌ی ۲.۲ برای یک $\mu > ۰$. تعریف کنید $r_۰ = b - Ax^{(۰)}$ و $h = x - x^{(۰)}$. سپس مسئله‌ی ۲.۲ به شکل زیر می‌شود.

$$min_{h\in R^n} \left\{||Ah - r_۰||^۲ + \mu||Lh||^۲\right\} \qquad (۳.۲)$$

بنابراین، h تقریبی از خطای $x_{true} - x_۰$ را ارائه می‌دهد. تقریب بهبود یافته x_{true} با $x^{(۱)} = x^{(۰)} + h$ داده می شود. تکرار این الگوریتم، روش Iterated Tikhonov regularization خواهد بود.

1. Compute $r^{(k)} = b - Ax^{(k)}$
2. Solve
$$min_{h\in R^n} \left\{||Ah - r_k||^2 + \mu^{(k)}||Lh||^2\right\}$$
to obtain $h^{(k)}$
3. Update $x^{(k+1)} = x^{(k)} + h^{(k)}$

هر ایتریشن این الگوریتم را می‌توان به صورت حل مسئله‌ی زیر مشخص کرد.

$$x^{(k+۱)} = x^{(k)} + (A^T A + \mu^{(k)} L^T L) - A^T(b - Ax^{(k)}) \qquad k = ۰, ۱, \ldots \qquad (۴.۲)$$

تجزیه‌ی GSVD جفت ماتریس $\{A, L\}$ به صورت زیر خواهد بود.

$$A = U\Sigma Y^T \qquad L = V\Lambda Y^T$$

که ماتریس‌های $U \in R^{m\times m}$ و $V \in R^{p\times p}$ متعامد هستند و ماتریس $Y \in R^{n\times n}$ یک ماتریس nonsingular است. همچنین داریم

$$\Sigma = diag(\alpha_۱, \alpha_۲, \ldots, \alpha_p, ۱, ۱, \ldots, ۱) \in R^{m\times n}$$
$$\Lambda = (diag(\lambda_۱, \lambda_۲, \ldots, \lambda_p), ۰, ۰, \ldots, ۰) \in R^{p\times n}$$

p المان اول Σ و Λ به نحوی مرتب شده‌اند که داریم

$$0 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_p \leq 1$$
$$1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$$
$$\alpha_i^2 + \lambda_i^2 = 1 \qquad 1 \leq i \leq p$$

پس مسئله‌ی ۴.۲ به صورت زیر تبدیل می‌شود.

$$(A^TA + \mu^{(k)}L^TL)x^{(k+1)} = A^Tb + \mu^{(k)}L^TLx^{(k)}$$
$$\left(Y\Sigma^T\Sigma Y^T + \mu^{(k)}Y\Lambda^T\Lambda Y^T\right)x^{(k+1)} = Y\Sigma^TU^Tb + \mu^{(k)}Y\Lambda^T\Lambda Y^Tx^{(k)}$$

از آن جایی که ماتریس Y nonsingular است می‌توانیم دو طرف معادله را از چپ در Y^{-1} ضرب کنیم. با گرفتن $z^{(k)} = Y^Tx^{(k)}$ داریم.

$$(\Sigma^T\Sigma Y^T + \mu^{(k)}\Lambda^T\Lambda)z^{(k+1)} = \Sigma^TU^Tb + \mu^{(k)}\Lambda^T\Lambda z^{(k)}$$

$$z^{(k+1)} = (\Sigma^T\Sigma Y^T + \mu^{(k)}\Lambda^T\Lambda)^{-1}(\Sigma^TU^Tb + \mu^{(k)}\Lambda^T\Lambda z^{(k)}) \tag{5.2}$$

از آن جایی که دو ماتریس Σ و Λ قطری هستند حل مسئله ۵.۲ آسان است.

۴.۲ The discrepancy principle and zero-finders

انتخاب پارامتر رگولاریزیشن مهم است و در نتیجه تاثیر دارد. مقدار پارامتر رگولاریزیشن $\mu^{(k)}$ به‌گونه‌ای تعیین می‌شود که راه‌حل تقریبی $x^{(k+1)}$ شرط زیر را برآورده کند.

$$||Ax^{(k+1)} - b||^2 = (\eta\delta)^2 \tag{6.2}$$

این شرط را بیشتر بررسی می‌کنیم. پاسخ مسئله در ایتریشن k ام به صورت زیر است.

$$x^{(k+1)} = (A^TA + \mu^{(k)}L^TL)^{-1}(A^Tb + \mu^{(k)}L^TLx^{(k)})$$

این جواب را در معادله‌ی ۶.۲ قرار می‌دهیم. از تجزیه‌ی GSVD و $\hat{b} = U^Tb$ و $z^{(k)} = Y^Tx^{(k)}$ داریم:

$$||Ax^{(k+1)} - b||^2 = ||\Sigma(\Sigma^T\Sigma + \mu^{(k)}\Lambda^T\Lambda)^{-1}(\Sigma^T\hat{b} + \mu^{(k)}\Lambda^T\Lambda z^{(k)}) - \hat{b}||^2$$
$$\geq \sum_{i=1}^{p}\left(\frac{\hat{b_j}\lambda_j^2 - \alpha_j z_j^{(k)}\lambda_j^2}{\frac{1}{\mu^{(k)}}\alpha_j^2 + \lambda_j^2}\right)^2$$

نامساوی به این خاطر است که فقط p ترم اول را در نظر گرفتیم و ترم‌های بعد از p را در نظر نگرفتیم. با تغییر متغیر $\beta = 1/\mu^{(k)}$ داریم.

$$\Phi(\beta) = \sum_{i=1}^{p}\left(\frac{\hat{b_j}\lambda_j^2 - \alpha_j z_j^{(k)}\lambda_j^2}{\beta\alpha_j^2 + \lambda_j^2}\right)^2 \leq ||Ax^{(k+1)} - b||^2$$

پس قصد داریم معادله زیر را حل کنیم.

$$\Phi(\beta) = (\eta\delta)^{۲} \tag{۷.۲}$$

مشتق اول و دوم این تابع به صورت زیر است.

$$\Phi'(\beta) = \sum_{i=۱}^{p} \frac{-۲\alpha_j^{۲}\left(\hat{b}_j\lambda_j^{۲} - \alpha_j z_j^{(k)}\lambda_j^{۲}\right)^{۲}}{\left(\beta\alpha_j^{۲} + \lambda_j^{۲}\right)^{۳}}$$
$$\Phi''(\beta) = \sum_{i=۱}^{p} \frac{۶\alpha_j^{۴}\left(\hat{b}_j\lambda_j^{۲} - \alpha_j z_j^{(k)}\lambda_j^{۲}\right)^{۲}}{\left(\beta\alpha_j^{۲} + \lambda_j^{۲}\right)^{۴}}$$

به طور کلی، $\Phi'(\beta) < ۰$ و $\Phi''(\beta) > ۰$ برای $\beta \geq ۰$. نتیجه می شود که تابع $\Phi(\beta)$ به طور یکنواخت کاهشی و محدب است. الگوریتم حل به شکل زیر خواهد بود.

Algorithm 1. $[x,\mu] = \mathrm{ITGSVD}(A, L, b, \eta\delta)$

Compute the GSVD $A = U\Sigma Y^T$ and $L = V\Lambda Y^T$
$\beta^{(0)} = 0,\ x^{(-1)} = x^{(0)} = 0,\ z^{(0)} = 0$
$\hat{b} = U^T b,\ k = 0$
while $\|x^{(k)} - x^{(k-1)}\| \geq \eta\delta\|x^{(k-1)}\|$ **do**
  Solve $\Psi_i(\beta) = 0$ for $\beta^{(k)}$. Ψ_i depends on $\eta\delta$ and is defined in the text.
  $z^{(k+1)} = (\Sigma^T\Sigma + \frac{1}{\beta^{(k)}}\Lambda^T\Lambda)^{-1}(\Sigma^T\hat{b} + \frac{1}{\beta^{(k)}}\Lambda^T\Lambda z^{(k)})$
  $k = k + 1$
end while
$x = Y^{-T} z^{(k+1)}$
$\mu = 1/\beta^{(k)}$

شکل ۱.۲: الگوریتم ITGSVD

زمانی که دو تکرار متوالی به اندازه کافی نزدیک باشند، تکرارها پایان می‌یابند. ما سه zero-finder را برای حل $\Phi(\beta) = (\eta\delta)^{۲}$ مقایسه می کنیم.

۱. روش نیوتن برای تعیین صفر

$$\psi_{۱}(\beta) = \Phi(\beta) - (\eta\delta)^{۲}$$

۲. روش نیوتن برای تعیین صفر

$$\psi_{۲}(\beta) = \frac{۱}{\Phi(\beta)} - \frac{۱}{(\eta\delta)^{۲}}$$

۳. یک صفر یاب که به صورت مکعبی همگرا می‌شود که روی تابع $\psi_{۱}(\beta)$ اعمال می‌شود.

۵.۲ Numerical examples

در این مسائل در صورت مشخص نشدن از $\eta = ۱/۰۱$ و $x^{(۰)} = ۰$ و ماتریس رگولاریزیشن سه قطری زیر استفاده می‌کنیم.

$$\begin{bmatrix} -۱ & ۲ & -۱ & & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & -۱ & ۲ & -۱ \end{bmatrix} \in R^{(n-۲)\times n}$$

که یک صورت گسسته اسکیل شده از عملگر مشتق دوم است. بردار خطا $e \in R^n$ با نویز گاوسی سفید مدل شده است و به نسبت زیر سطح نویز اشاره می‌کنیم.

$$\sigma = \frac{||e||}{||b_{true}||}$$

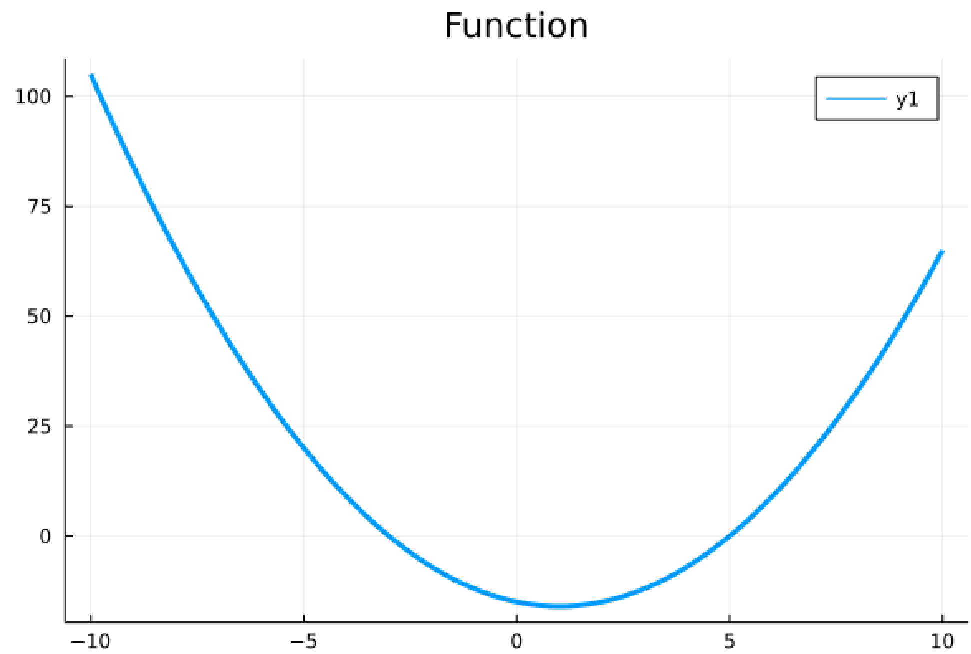
خطای نسبی بازسازی $x_{\mu^{(k)}}$ با این معادله به دست می‌آید.

$$RRE(x_{\mu^{(k)}}) = \frac{||x_{\mu^{(k)}} - x_{true}||}{x_{true}}$$

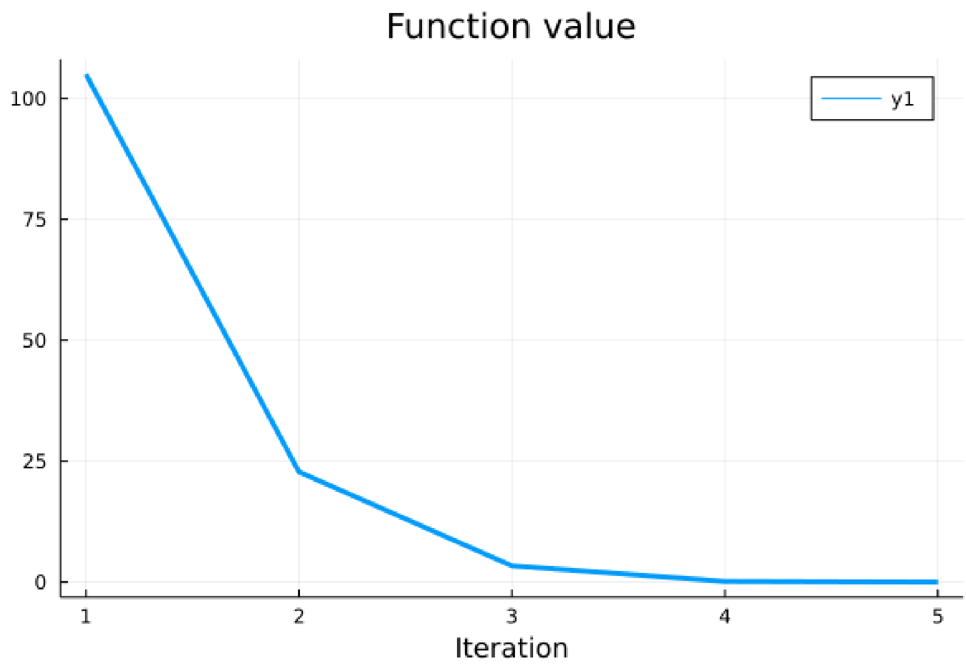
برای حل این مسئله به zero-finder ها نیاز داریم. روش نیوتن را در این قسمت روی تابع $f(x) = x^۲ - ۲x - ۱۵$ اعمال می‌کنیم. روش نیوتن به شکل زیر است که در هر ایتریشب سعی می‌کند به سمت ریشه‌ی تابع حرکت کند. صورت دقیق روش نیوتن به شکل زیر است.

$$x_{k+۱} = x_k - \frac{f(x_k)}{f'(x_k)}$$

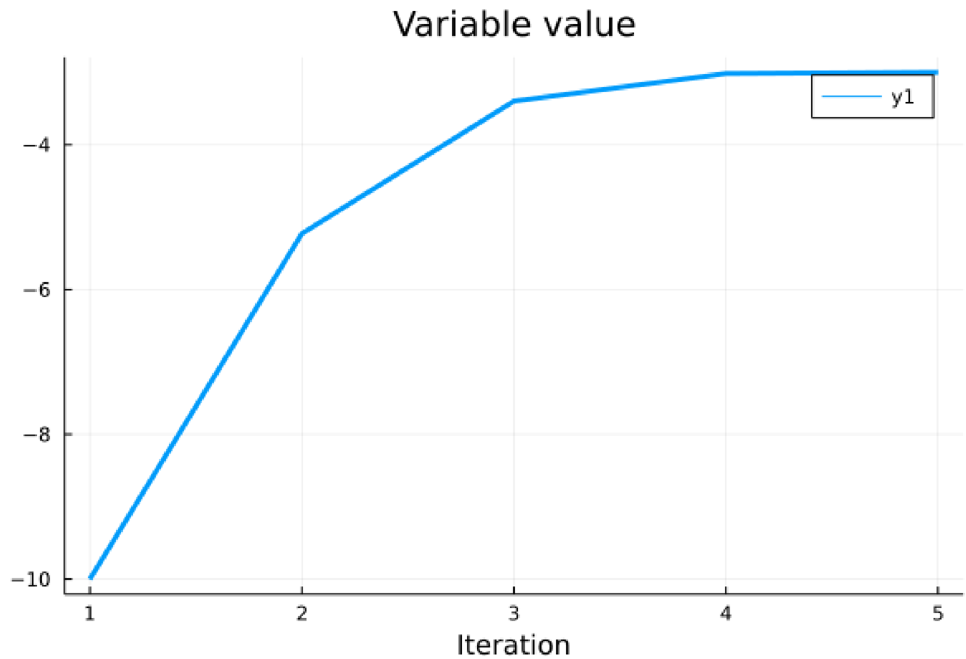
حال در کد Newton Method این دو روش را پیاده سازی کردم. نتایج به صورت زیر است.



شکل ۲.۲: تابع f(x)



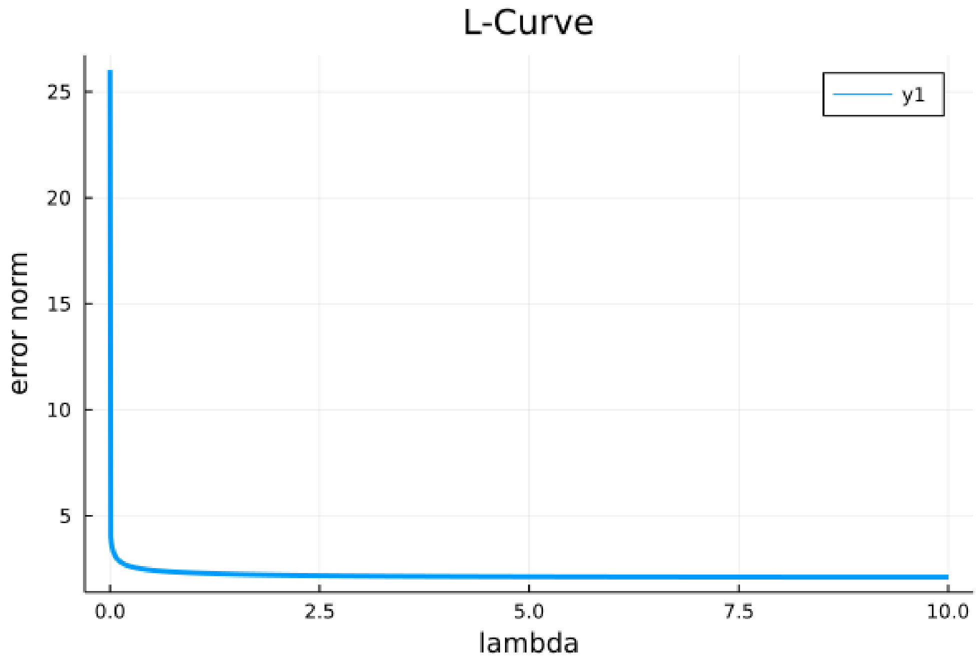
شکل ۳.۲: تغییرات تابع هدف به ازای ایتریشن‌ها



شکل ۴.۲: تغییرات متغیر به ازای ایتریشن‌ها

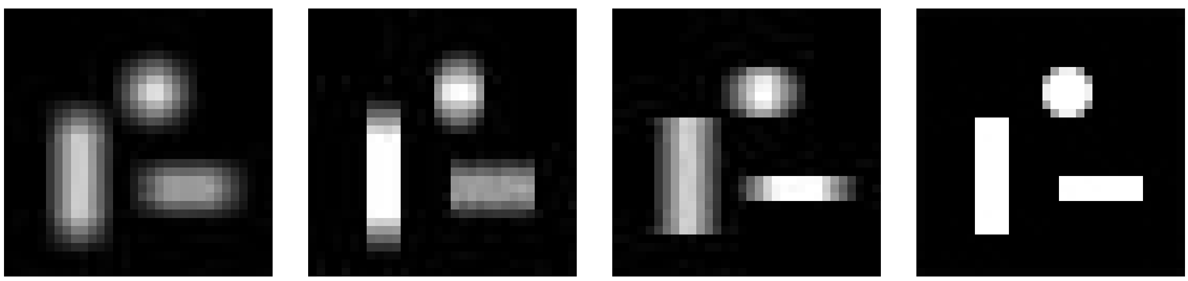
همان طور که مشخص است با ۵ بار تکرار الگوریتم از نقطه‌ی ۱۰− به نزدیک‌ترین ریشه رسیدیم. از آن جایی که تابع ما محدب است، تنها یک ریشه در صورت ریشه داشتن خواهد داشت.
برای حل مسئله به صورت اولیه یک ماتریس رندوم $A^{n\times n}$ تولید کردم. مقدار ویژه‌هایش را اعداد رندوم در بازه‌ی [۰, ۰/۱] قرار دادم. ماتریس b را هم به صورت رندوم انتخاب کردم. نویز گوسی سفید تولید کردیم و مسئله را در نظر گرفتم. برای یک حالت خاص، در صورتی که از روش معمول pseudo-inverse استفاده کنیم نرم جواب به دست آمده

و جواب درست برابر ۲۶/۰۱ خواهد بود. مسئله را برای حالت‌های مختلف λ حل کردم و L-curve را به صورت زیر به دست آوردم.



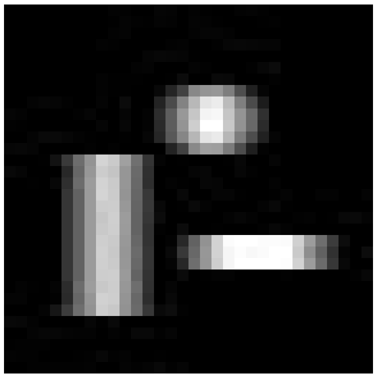
شکل ۵.۲: L-curve

همان طور که مشخص این منحنی اهمیت این رگولاریزشن را مشخص می‌کند.
برای بررسی یک مورد دیگر مسئله‌ی حل شده در مقاله را در نظر می‌گیریم. در این مسئله سعی می‌کنیم عکس blur شده را واضح کنیم. عکس‌های ورودی در سایز [۳۲ × ۳۲] وجود دارد. هر عکس به بردار ستونی تبدیل می‌شود و با یک ماتریس $A^{۱۰۲۴\times۱۰۲۴}$ به ورژن مات شده اش نگاشت می‌کند. ۳ تبدیل در زبان متلب برای blur کردن وجود داشت که نام آن‌ها را به ترتیب A_x A_y و T قرار دادم. . یک تصویر ورودی تولید کردم و ورژن مات شده اش را به دست آوردم که به شکل زیر شد.

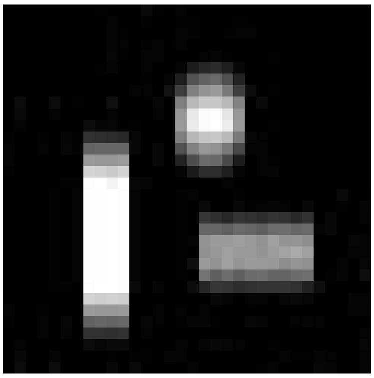


(الف) تصویر اولیه (ب) تصویر مات شده با A_x (ج) تصویر مات شده با A_y (د) تصویر مات شده با T

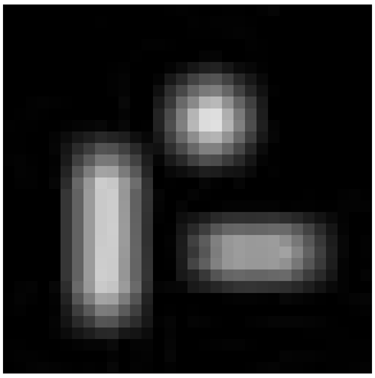
تصاویر و ماتریس‌ها را در متلب ذخیره کردم و به جولیا فرستادم. حال در جولیا می‌خواهیم مسئله‌ی کمترین مربعات را برای پیدا کردن عکس اصلی از بین عکس‌های بلور شده پیاده سازی کنیم. الگوریتم ۱.۲ را پیاده‌سازی کردم. در قسمت توابع $\Phi(\beta)$ و $\Phi'(\beta)$ را نوشتم و با استفاده از روش نیوتن مسئله را حل کردم. مقدار اولیه‌های مورد استفاده منطبق با الگوریتم ذکر شده است. نتیجه‌ی این الگوریتم به شکل زیر شد.



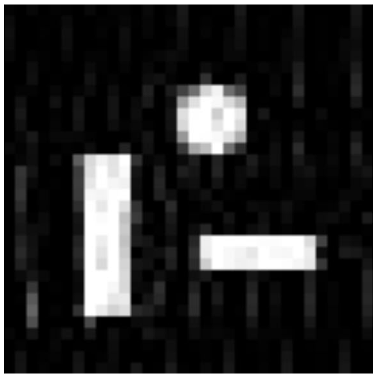
(الف) تصویر مات شده با A_x



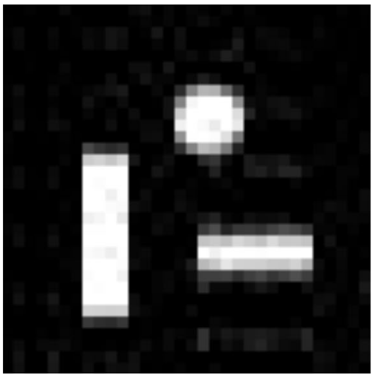
(ب) تصویر مات شده با A_y



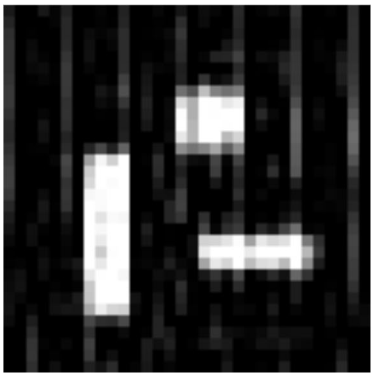
(ج) تصویر مات شده با T



(د) تصویر بازسازی شده‌ی متناظر با A_x



(ه) تصویر بازسازی شده‌ی متناظر با A_y



(و) تصویر بازسازی شده‌ی متناظر با T

هیچ کدام از تبدیل‌های استفاده شده وارون پذیر هم نبودند. نتایج به دست آمده نشان می‌دهد که شکل تا جای ممکن بازسازی شده‌است.

کد این سوالات در فایل‌های deblurring A_x، deblurring A_y و deblurring T قرار دارد.