Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Efficient COVID-19 testing via contextual model based compressive sensing

Mehdi Hasaninasab, Mohammad Khansari

Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

ABSTRACT

The COVID-19 pandemic is threatening billions of people’s life all over the world. As of March 6, 2021, COVID-19 has confirmed in 115,653,459 people worldwide. It has also a devastating effect on businesses and social activities. Since there is still no definite cure for this disease, extensive testing is the most critical issue to determine the trend of illness, appropriate medical treatment, and make social distancing policies. Besides, testing more people in a shorter time helps to contain the contagion. The PCR-based methods are the most popular tests which take about an hour to make the output result. Obviously, it makes the number of tests highly limited and consequently, hurts the efficiency of pandemic control. In this paper, we propose a new approach to identify affected individuals with a considerably reduced No. of tests. Intuitively, saving time and resources is the main advantage of our approach. We use contextual information to make a graph-based model to be used in model-based compressive sensing (CS). Our proposed model makes the testing with fewer tests required compared to traditional testing methods and even group testing. We embed contextual information such as age, underlying disease, symptoms (i.e. cough, fever, fatigue, loss of consciousness), and social contacts into a graph-based model. This model is used in model-based CS to minimize the required test. We take advantage of Discrete Graph Signal Processing on Graph (DSP_G) to generate the model. Our contextual model makes CS more efficient in both the number of samples and the recovery quality. Moreover, it can be applied in the case that group testing is not applicable due to its severe dependency on sparsity. Experimental results show that the overall testing speed (individuals per test ratio) increases more than 15 times compared to the individual testing with the error of less than 5% which is dramatically lower than that of traditional compressive sensing.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

COVID-19 has become the main threat to people’s lives all over the world. It also disrupts both business and social activities. Economic depression is another problem that all people are involved in, regardless of affected or non-affected. In the absence of a vaccine and a definite cure, the only remedy is that stop spreading the viruses. Large-scale quarantine is not a good solution because of serious, perhaps irreparable consequences in long term including a downturn and social problem. The best policy which minimizes economical damage is to isolate the infected people. So, COVID-19 testing is the primary step to isolate infected people. This testing must be fast enough to cover most people in a short time and consequently isolate the right individuals before time passed. There are mainly two types of test that can detect COVID-19 in individuallys: 1) Serological tests, which study the presence of antibodies in the blood of individuals, and 2) The Swap test, which takes material from the cavity between nose and mouth and looks for RNA of an alive virus. Although the serological test has its advantages, the swap test is extremely recommended by the Center for Disease Control and Prevention (CDC) [1]. This class of tests uses Real-Time Polymerase Chain Reaction (RT-PCR) for detecting COVID-19. This method is based on selective DNA strands amplification in each cycle which can make the result in a shorter time and conservation of reagent and testing kits. Additionally, increasing the range of detection makes it suitable techniques to detect COVID-19. The RT-PCR is done typically in 40 cycles, and in each cycle, three major steps are performed [2]:

1) Denaturation

High temperature (typically 95 °C, the maximum temperature that DNA can withstand) incubation is used to “melt” double-strand DNA to the single stand.

1) Annealing

https://doi.org/10.1016/j.patcog.2021.108253
0031-3203/© 2021 Elsevier Ltd. All rights reserved.
During annealing, the temperature is reduced (usually 5 °C below the first step) to let the complementary sequence hybridize.

1) Extension

At 70–72 °C, the optimal activity of DNA polymerase is achieved and primer extension occurs.

It is expected to take 8 h to complete the RT-PCR test for each individual. This is a long time that extremely limits the number of individual testing per day.

Although the RT-PCR test is recognized as the golden standard for determining COVID-19 status in individuals [3], but it’s time-consuming is the critical challenge. It may cause to diagnose from chest CT image as an alternative way [4].

One of the usual methods to mitigate this limitation is group testing [5]. Group testing is an old technique introduced by Dorfman in 1943. In the group testing, people are grouping and testing is done on the pooled sample. If the result is negative, it means all group members are disease-free; otherwise, the test must be performed for each group member separately. Although it has advantages compared to sampling techniques (i.e. no data loss), it has two important restrictions: The output of the group testing is a binary value. In other words, it just determines disease affection or non-affection. This is inefficient in the case of COVID-19 where the viral load (i.e. virus concentration) is important to determine the COVID-19 severity among patients [6] and also to study about the antibody response [7].

The other drawback of group testing is its dependency on the size of groups. Increasing the group size increases the number of required test and fades the efficiency of the group testing [5]. This is extremely critical in case of a large population (e.g. testing COVID-19).

From the last decade, Compressive sensing (CS) has been a popular technique in signal processing related fields. The key advantage of CS over traditional sampling is the extremely lower sampling rate required for the signal recovery. CS has two main phases: signal sampling and signal recovery. Contrary to the traditional sampling, the CS samples cannot directly be mapped to any specific data elements, but a linear combination of all data elements. In each CS sample, each individual's data have its own weight which is specified by the sampling matrix. In the recovery phase, the signal elements are recovered from CS samples. From the mathematical point of view, CS signal recovery is going to find a solution (i.e. the original signal) of the underdetermined system of equations that have an infinite number of solutions [8]. It is done by considering the signal's unique features (i.e. sparsity) among other solutions.

In the application of COVID-19 testing, the CS sampling rate is the number of required tests to identify all group members' COVID-19 affection status (i.e. recovered signal). Depending on the recovery strategy chosen, the appropriate sampling rate may be varied [9]. Even though CS efficiency, its implementation has some drawbacks in the real world. In the COVID-19 testing scenario, decreasing the number of required tests while maintaining the required quality is the most critical challenge. Baraniuk introduced the model-based CS to present the unique features concretely in the form of a mathematical model. In model-based CS, in addition to sparsity, the structure of signal values and locations are used as signal features. It has been shown that model-based CS makes it possible to reduce the sampling rate without sacrificing robustness and the quality of the output [10].

To the best of our knowledge, the model-based CS hasn’t applied to COVID-19 so far. Moreover, in none of the proposed models, the interdependency between signal elements (i.e. the relation between individual test results in the context of COVID-19) was considered. In this paper, we introduce the Contextual Model-based CS (CMb-CS) that considers the contextual information (e.g. lifestyle, underlying disease, and social contacts) to determine the signal elements (i.e. individual test results) interdependency. The model is generated based on the emerging field of Signal Processing on Graph (DSPG) and exploit its powerful features to show interdependency between signal elements. The aim of our model is to represent the signal characteristic to recover the compressed signal (i.e. compressed COVID-19 test results) more efficiently. The interdependency between signal element, along with the signal smoothness and sparsity are three factors that are embedded into our model to represent the signal more precisely. Owing the clear image that the model offers from the recovered signal, there are expected that the signal is recovered with much lower number of data. At the final part of this paper, the model efficiency in COVID-19 testing is compared to the existing group testing and the CS based method.

The rest of this paper is organized as follows: Section 2 describes some necessary backgrounds. In Section 3, the proposed model is presented. In section 4 the experimental results are shown and explained, and finally, the conclusion is discussed in section 5.

2. Backgrounds

2.1. Group testing

Group testing was firstly introduced by Dorfman to detect syphilis from blood samples in a large population without the need to test each individual [5]. Instead of testing each individual separately, the population of N people is divided into groups of n members. In each group, the samples are mixed and the test is done on the pooled samples. The negative result means all group members are free from infection. Otherwise, there is at least one person contaminated. It is proven that in this way, the number of tests per individual is reduced. The efficiency of group testing depends on two key parameters: infection rate (e.g. probability of infection) and group size (n). The expected relative cost (C) given by:

\[ C = \frac{n+1}{n} - (1-P)^n \]  

(1)

Where n is the size of the group (number of members), and P is the average infection probability.

Group testing has been applied on COVID-19 testing[11]. The multistage group testing approach was proposed to reduce the number of COVID-19 required tests. In this work, the group testing was done iteratively, and in each iteration, the group size was changed dynamically. It was shown that this approach reduce the above C parameter [11]. From (1), it is clear that increasing n or P may causes C > 1, which means that using group testing is ineffective. Therefore, generally group testing is not beneficial in a large population. Moreover, it gives only binary results (infection or non-infection) and the viral load is also unspecified. These shortcomings make the group testing not a suitable method for COVID-19 pandemic conditions (i.e. high population, and fast result required, and high output quality demanding). Another demand is to reduce more number of required tests which consequently, saves quantity of reagents and manpower. Regarding all above together, one of the promising approaches to address these challenges, is compressive sensing which is described in the following section.

2.2. Compressive sensing

CS is an emerging field of research that merges sampling and compression in one step. Contrary to traditional sampling, in the CS, it has been proven that the signal can recover completely at a much lower sampling rate than that in Shannon/Nyquist [8]. Due to the merging of data acquisition and data compression in one
single step, there will be no need to access all data elements. One of the key concepts in CS is sparsity which is described in the following.

1) Sparsity

Imagine the $x \in \mathbb{R}^N$ is a vector consist of the N individual test result. Vector $x$ is K-sparse if it has at most K nonzero coefficients ($K < < N$).

1) Sensing/Measurement Matrix

The first step for applying the CS to a signal is the sensing process for which the sensing operator $\Phi^{M \times N}$ ($M < < N$) is used. The compressive measurement is shown as:

$$ y = \Phi x \quad (2) $$

1) CS recovery

Where $y \in \mathbb{R}^M$ is the measurement (i.e. sensed) vector. Clearly, each CS sample is a linear combination of all signal elements. The CS technique is a suitable alternative to the Shannon/Nyquist based sampling scheme if the signal is sparse. It can be easily shown that the CS signal recovery problem is solving an underdetermined linear equation system (i.e. the number of unknowns is more than the number of equations). Given that there are an infinite number of solutions, it is necessary to use prior knowledge of the signal to choose the most appropriate solution ($x$). The sparsity level is one of the key knowledge that is exploited in the traditional CS recovery.

Suppose that $x$ is a K-sparse signal. The CS recovery process is finding the signal which satisfies two conditions: first, $x$ is K-sparse, and second, $x$ satisfies (2). Suppose also that $x$ has N elements, so finding a K-sparse solution simply means searching for a solution among $\binom{N}{K}$ different signals (3).

$$ \text{minimize} \| x \|_0 \quad \text{subject to:} \quad y = \Phi x \quad (3) $$

Where $\| x \|_0$ is the $x$ zero norm which is the number of nonzero elements of $x$.

While the CS shows significant improvement from traditional sampling in terms of both sampling rate and output quality, it has some drawbacks. Computing $x$ from $\binom{N}{K}$ possible solutions is a time-consuming process, because of the large search space. One possible way to improve the reconstruction process is to limit the search space. It is proven that in this way the reconstruction time is decreased. Furthermore, reconstruction can be done at a much lower sampling rate without sacrificing signal quality [10] (i.e. individual test results).

Limiting search space could be achieved using additional knowledge beyond sparsity. This is the core idea that led to the emergence of the model-based CS, which is a suitable alternative to traditional CS, especially in time and sample constraint applications such as COVID-19 testing. In this way, due to the reduction of the search space too much lower than $\binom{N}{K}$, the recovery process can be done with even fewer CS measurements. It means that with a lower number of tests, more people COVID-19 status can be identified.

2.3. Discrete signal processing on graph (DSC)

Graph signal processing is an emerging field that tries to express the regular and irregular structured signals [12]. Social networks and wireless sensor networks are examples of those signals. The signal structure can be represented by a graph form appropriately; the edge weight between two nodes is proportional to the similarity or correlation of them. This unique feature makes the graph a powerful tool to represent the interrelationship between signal elements. In the COVID-19 testing application, the first step is to model the data in the DSPC domain. In the following subsections, the graph signal model and also some of the key concepts in DSPc are discussed.

1) Graph signal model in COVID-19 testing scenario

Consider a graph $G(V, W)$ where $V = \{ v_0, v_1, \ldots, v_{N-1} \}$ is graph node set and $X = \{ x_0, x_1, \ldots, x_{N-1} \}$ is a signal value in each node. In graph signal representation, the signal elements can be represented by nodes value and the relations between these elements are represented by graph weight. In the COVID-19 testing scenario, each node corresponds to an individual, $X$ is an output vector and its elements $x_i$ is the health status of $i^{th}$ individual. $W$ is the weighted adjacency matrix and $W_{ij}$ is the weight between the nodes $i$ and $j$ (i.e. individual $i$ and $j$). It is clear that for the undirected graph, $W$ is a symmetric matrix (i.e. $W_{ij}=W_{ji}$) and $W_{ii}=0$ if there is no correlation between the nodes $i$ and $j$. In COVID-19 testing scenario, the weighted adjacency matrix is an appropriate tool to represent the degree of similarity between individuals, which is a part of the proposed model. Fig. 1 shows the weighting advantage of DSPc. In the traditional form of signal representation, the relation between signal elements (i.e. red line) is not considered. So all the three signals $G_1, G_2, G_3$ are the same in traditional signal representation, but they differ in DSPc domain.

1) Graph parameters

The adjacency matrix of the graph $G$ is defined as $A_{m \times n} = (A_{ij})$ in which:

$$ A_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \quad (4) $$

The 1’s in the adjacency matrix can be replaced with $W_{ij}$ which typically is the correlation between $v_i$ and $v_j$.

1) Graph eigenvalue and eigenvector

For graph G with adjacency matrix $A \in \mathbb{R}^{m \times n}$, for the vector $\lambda \neq 0$ the eigenvalue $\lambda$ is defined as follows:

$$ AV = \lambda V \quad (5) $$

The vector $x$ is an eigenvector corresponding to $\lambda$.

1) Graph shift

The graph shift is one of the basic operations in DSPc. Graph shift simply replaces the signal value at the node $v_i$ with the linear combination of signal values in neighborhood nodes.

$$ \tilde{x}_i = \sum_{m=0}^{N-1} A_{m,n} x_m \quad (6) $$

1) Graph total variation

In traditional signal processing, the Total Variation (TV) is the difference between two consecutive samples [12]. In the case of the COVID-19 testing scenario, TV has a conceptual meaning which shows how the graph signal of test results varies. The TV measures the difference between each node and its neighbors on the graph. In DSPc, TV is generally defined as a similarity between the graph signal and its shifted version. In fact TV is derived from the local variation which somehow represents the counterpart of the signal gradient in traditional signal processing [13]. Graph variation is a highly popular parameter in optimization and learning problems. It characterizes the smoothness of the graph, which is the key parameter to find the discriminative graph signal. Using the fact that discriminative signals have different values in multiple groups, low
smoothness value can be one of the features in that type of graph signal. It can also be used as the key constraint in learning related problems including graph based learning methods [14]. As in [15], the edge weight smoothness and graph sparsity is considered as two main criteria for dimension reduction and semi-supervised classification. In DSPC, the local variation (gradient) at the $n$th node is defined as:

$$S(n) = x_n - \bar{x}_n$$  \hspace{1cm} (7)

Substitute (6) in (7) yields:

$$S(n) = x_n - \sum_{m=0}^{N-1} A_{m,n}^{\text{norm}} x_m$$  \hspace{1cm} (8)

Where $A^{\text{norm}}$ is a normalized version of $A$:

$$A^{\text{norm}} = \frac{1}{|\lambda_{\text{max}}(A)|} A$$  \hspace{1cm} (9)

Where $\lambda_{\text{max}}(A)$ is the largest eigenvalue of $A$. Using $A^{\text{norm}}$ in (8) guarantees the graph signal is scaled properly compared with the original one.

To make sure that the shifted graph signal is scaled properly compared to the original one, $S(n)$ can be rewritten as:

$$S(n) = x_n - \sum_{m=0}^{N-1} \frac{1}{|\lambda_{\text{max}}(A)|} A x_m$$  \hspace{1cm} (10)

The magnitude of local variation of the graph in each node is denoted by $S(n)$ [13]. The total variation is naturally the sum of local variations in each node. TV is represented based on $\ell_p$-norm:

$$S_p(X) = \left\| X - \frac{1}{|\lambda_{\text{max}}(A)|} AX \right\|_p$$  \hspace{1cm} (11)

For $p = 2$, it is called the quadratic form of graph total variation [16].

3. The proposed method

Let us now introduce the CMb-CS for COVID-19 testing. The proposed method can be divided into two main phases: sampling and recovery.

The schematic of the sampling phase is depicted in Fig. 2. In the CS sampling phase, the samples are pooled according to the weights specified in the sampling matrix. The pooled sample is fed into the RT-PCR machine. So far, it is just similar to the group testing except that the weighting mechanism. The weight values in group testing are the binary values (0 or 1). Although the binary weighting matrix is also applicable in CS-based crowd sensing applications, some of the sensed elements are not contributed in the CS measurement, which may lead to a less reliable result in the recovered signal [17].

In the proposed method, instead of the binary value, the sampling matrix elements has continuous values in $[-1,1]$.

At the recovery phase, the aim is to recover the signal from CS samples. As mentioned before, in the model-based CS, the main goal is to limit the search space by applying a model to find the solution more efficiently. Our proposed model employs three limiting factors including graph signal total variation (TV), node value sparsity (as a key factor for limiting CS recovery search space), and the structural similarity between the graph of recovered results and the reference graph which exploit contextual information (See Fig. 3). The TV and the sparsity have been introduced in the background section. Regarding the graph model, this model is used to represent the contextual data in the CS recovery process. In addition to infection status, for each individual, some data can be extracted from individual health records. This contextual data contains age, underlying disease, symptoms (i.e. cough, fever, fatigue, loss of consciousness), and contacts to infected individuals. This contextual information is exploited in the graph model which is the third component of our model based CS recovery process. The schematic of this graph is depicted in Fig. 4.

As can be seen, the graph is a full mesh graph and each node and the weights correspond to the individuals and the interdependency between individuals’ state respectively. As mentioned in the background section, the weights represent the similarity between the contextual information:

$$W_{ij} = \frac{1}{7} \left( 1 - \frac{|\text{Age}_i - \text{Age}_j|}{\text{max}\{\text{Age}_i, \text{Age}_j\}} + \text{Ci}(i,j) + \text{Fev}(i,j) + \text{Co}(i,j) + \text{Fa}(i,j) + \text{SoB}(i,j) + \text{LoC}(i,j) \right)$$  \hspace{1cm} (12)

Where:

- **Ci**: Contact with an infected individual
- **Fev**: Fever
- **Co**: Cough
- **Fa**: Fatigue
- **SoB**: Shortness of Breath
- **LoC**: Loss of Consciousness

Expect than age, all the mentioned parameters take binary values. Each parameter represent the similarity between two individuals in the specific contextual information. For example, $\text{Fev}(i,j) = 1$ if both individual $i$ and $j$ have fever, otherwise $(i,j) = 0$.

This $W_{ij}$ appeared in the model as the reference graph edge weights. The similarity between the reference graph and recovered result graph (as depicted in Fig. 3), along with TV and sparsity, are the limiting factors for the CS search pace. For CS COVID-19 test result recovery, finding the solution in limited search space is easier and can be done with considerably fewer sample.

We now formulate the recovery problem in DSPC. The basis for test results recovery is expressed in (13). To recover the test results for each individual $(x)$ from the CS measurement data $y$, $x$ must be a solution for (2) which typically has an infinite number of solutions [9]. To select an individual test results from this infinite number of solutions we must solve a combinatorial multi-parameters optimization problem. In the proposed method, the proper solution (i.e. the individual test result) must minimize the three mentioned
limiting factors of the search space and also satisfy (2) at the same time:

\[
\hat{X} = \text{arg min } \alpha S_p(X) + \beta \|X\|_q + \gamma \|A^{\text{temp}} - A\|_r\]

subject to \( y = \Phi X \) \hspace{1cm} (13)

Where, \( \hat{X} \) is the recovered signal (i.e. COVID-19 testing result vector), \( S_p(X) \) is the \( p \)-norm of graph total variation, \( X_0 \) is the graph \( \ell_q \)-norm node value (e.g. node value sparsity), and \( A^{\text{temp}} - A \) measures the structural similarity between the recovered signal graph and contextual information graph. This measurement is done simply by making the \( \ell_r \)-norm difference between the contextual information graph adjacency matrix (\( A^{\text{temp}} \)) and the recovered data graph adjacency matrix (\( A \)). \( \alpha, \beta, \) and \( \gamma \) are the weights assigned to each term to adjust their effectiveness on the output result. It is proven that (13) is NP-hard in general and not easy to solve [16]. Setting proper values for \( q \) and \( r \) for each term can express (13) as a convex optimization problem. In the case of graph total variation \( S_p(X) \), although both \( \ell_1 \) norm and \( \ell_2 \) norm (i.e. \( p = 1 \) and \( p = 2 \), respectively) can tend to the convex form, here we use a quadratic norm (i.e. \( \ell_2 \)-norm) instead of an \( \ell_1 \)-norm because; first, it is computationally easier to optimize \( \ell_2 \)-norm than \( \ell_1 \)-norm based graph variation and second, the smooth graph signal can be more efficiently separated from non-smooth parts by \( \ell_2 \)-norm [16].

However, it is shown that in general \( K_q \) approaches \( K_0 \) when \( q \) tends to zero. Therefore, to relax the NP-hard problem, some other values of \( q \) and \( r \) are chosen for the second and third terms in (13), respectively. It is proven that for \( 0 < q < 1 \), the problem again becomes NP-hard. But for \( q = 1 \) it becomes a convex problem [18]. Also, by setting \( q = 2 \), the problem is still in convex optimization form but it does not provide a reasonable approximation to the original signal [8]. Similarly, \( r \) should set to 1. Here, the easier optimization process is the main advantage that motivates us to try to transform the problem into a convex optimization problem since the local minimum is also the global minimum. By setting \( p, q, \) and \( r \) to 2, 1, and 1 respectively in (13) and (14) is obtained as follows.

\[
\hat{X} = \text{arg min } \alpha S_2(\|X\|) + \beta X + \gamma \|A^{\text{temp}} - A\|_2\]

subject to \( y = \Phi X \) \hspace{1cm} (14)
As mentioned earlier, in this problem the recovery process is considered in three aspects: graph variation, graph sparsity, and graph structure similarity to contextual information graph. For a simple and efficient solution of (14), a class of decomposition-based methods is considered. These methods are the proper choices especially for solving large-scale statistical tasks. Another advantage is that this class of algorithms is flexible enough to solve multi-dimensional problems [19].

To solve (14), we apply the Alternating Direction Method of Multipliers (ADMM) in this case, to benefit from the advantages of both decomposability and fast convergence property [19]. Using the ADMM, the minimization problem can be decomposed to the individual sub-problems. As done in a recent work, the problem of minimizing the rank, sparsity, and total variation is solved by ADMM for principal component analysis for compressed sensing of dynamic MRI [20].

In ADMM, each variable updates iteratively until the convergence takes place (see Appendix 1). To solve (14) by employing ADMM, we rewrite it as follows.

$$\hat{X}^k = \arg \min \alpha S_2(X) + \beta \|X\|_0 + \gamma \|A^* - A\| + I(C_M)$$

Solving for X:

$$X = (\Phi^T \Phi)^{-1} \left( \frac{2a}{\eta \lambda_{\max}(A)} A - \frac{2a}{\eta} \Phi y + \Phi u^k \right)$$

(See Appendix 1 for details). To minimize (17), we can replace $S_2(X)$ by (11) and simply take derivation which yields:

$$a - \frac{a}{\lambda_{\max}(A)} A - \frac{\eta}{2} \Phi (y - \Phi X + u^k) = 0$$

Solve for $X$: 

$$X = (\Phi^T \Phi)^{-1} \left( \frac{2a}{\eta \lambda_{\max}(A)} A - \frac{2a}{\eta} \Phi y + \Phi u^k \right)$$

For $\|X\|_0$, we aim to solve:

$$\|X\|_{k+1} := \arg \min_F \left( \beta \|X\|_0 + \frac{\eta}{2} \|y - \Phi X_{k+1} + u^k\|^2_2 \right)$$

Similarity, for minimizing (20):

$$\beta - \frac{\eta}{2} \Phi X_{k+1}^T (y - \Phi \|X\|_0 X_{k+1} + u^k) = 0$$

Solving for $\|X\|_0$:

$$\|X\|_0 = \left( \frac{\eta}{2} \Phi^T \Phi \left( \frac{X_{k+1}}{\|X\|_0} \right)^T \frac{X_{k+1}}{\|X\|_0} \right)^{-1} \left( \frac{\eta}{2} \Phi X_{k+1} u^k + \frac{\eta}{2} \gamma \Phi X_{k+1} - \beta \right)$$

Where $X_{k+1} = \frac{X_{k+1}}{\|X\|_0}$ is the auxiliary variable.

And for $A$:

$$A_{k+1} := \arg \min \gamma \left( \|A^\text{temp} - A\| + \frac{\eta}{2} \|y - \Phi X_{k+1} + u^k\|^2_2 \right)$$

Also this follows the same as that for $\|X\|_0$ and $X$:

$$-\gamma A - \eta \Phi X_{k+1} + \frac{\eta}{2} \sum_{i=1}^N \text{degree}(v_i) + \|X_{k+1}\|^2 (y - \Phi X_{k+1} + u^k) = 0$$

For simplicity, set:

$$\sum_{i=1}^N \text{degree}(v_i) + \|X_{k+1}\| = P$$

Solving for $A$:

$$A = \eta \Phi \left( X_{k+1} + u^k \right) (\gamma \Phi X_{k+1} + u^k)^T \text{degree}(v_i) + \|X_{k+1}\|^2 (y - \Phi X_{k+1} + u^k)$$

Where:

$$u_{k+1} := u_k + (y - \Phi X_{k+1})$$

Defining $r^k = y - \Phi X_k$ and $u^k$ which is cumulative residual at kth iteration:

$$u^k = u^0 + \sum_{j=1}^k r^j$$

3.1. Stopping criterion

Eqs. (17) to (28) are solved iteratively until the two main conditions are met, i.e. the resultant signal satisfies (2), which is the core equation in compressive sensing, and the $XV$, $A$ remain nearly constant during the iteration. It means:

$$y - \Phi X^* \leq \varepsilon$$

Where $\varepsilon$ is the threshold for (2). Since ADMM converges slowly in cases that required high accuracy, we chose those thresholds small enough instead of absolute zero to make a tradeoff between signal quality and iteration number.

We obtained the closed-form solution by setting the derivative to zero. We just set $C_M = 0$ to satisfy the constraints.

For the final implementation, see Algorithm 1.
Algorithm 1: CMB-CS Recovery

Input: \( u^0 = 0, k = 0, X^0 = O^{n \times 1}, y^0 = O^{n \times 1}, \Phi^0 = O^{n \times h} \)

Output: \( X \)

While the stopping criterion is not satisfied

\[
X = (\Phi^T \Phi)^{-1} \left( \frac{1}{\sigma^2} A - \frac{2}{\gamma} \Phi y + \Phi u^k \right)
\]

\[
\|X\|_2 \leftarrow \left( \frac{1}{2} \|\Phi X^k\|_F \right)^{1/2} \left( \frac{1}{2} \|\Phi X^{k+1}\|_F + 2 \gamma \|X^{k+1}\|_F - \beta \right)
\]

\[
A \leftarrow \eta \Phi^T (X^{k+1} + u^k) \left( \frac{\gamma}{\gamma + \eta} \Phi + \Phi^T (\Phi \Phi^T)^{-1} \right) \left( X^{k+1} - X^{k+1} \right)
\]

End

\[
u^{k+1} = u^k + \gamma - \Phi X^{k+1}
\]

Do until the stopping criterion is met.

4. Experimental results

The proposed model is applied on data with the clinical characteristic of COVID-19 infected individuals in Iran [21]. This dataset contains 1024 individuals’ data which covers a wide variety of ages (15–85 years). In this dataset, the affection rate is 0.46 (i.e. 472 individuals were affected). In the case of group testing, the expected relative cost (C) in (1) is greater than one and so, the group testing is totally useless. The traditional CS and the proposed CMB-CS is applied on this dataset with various sampling rates ranging from 50 to 600 with the step size 10. The traditional CS and CMB-CS are compared in terms of Mean Square Error (MSE) (see Fig. 5), false positive, and false negative values. In order to calculate the false positive and the false negative values, we set a threshold. All entities below 0.5 are set to zero (i.e. negative result) and the rest of the entities set to one (i.e. positive result).

The results clearly show that applying our proposed model, with just 150 tests, can identify all 1024 individuals’ affection status correctly. While in traditional CS a similar result still can’t be achieved even by increasing fourfold the test numbers. It is due to the fact that traditional CS mainly relies on sparsity, which is not a highly discriminative feature in this case. In addition to the sparsity, our proposed model considers some real features (i.e. graph smoothness and graph similarity to the reference graph model). In Table 1, along with MSE, false positive and false negative results are represented at the various test numbers ranging from 50 to 200. As the table shows, our proposed method outperforms traditional CS in all number of tests.

4.1. The effect of contextual information

The effect of each contextual information item is represented in Fig. 6. It is clear that “age” and “loss of consciousness” have a major and minor effect on the result respectively. Comparing to the Traditional CS, it is clear that even if the CMB-CS uses only the least effective contextual information (i.e. “loss of consciousness”), it still outperforms traditional CS. Fig. 6 also reveals that the “age” and “fever” have considerable effect compare to the other contextual information and therefore, are the most dominant factors for identifying COVID-19. Regarding other contextual information, they are more separable from the sample rate of 260 onward.

4.2. The effect of model components

The sensitivity of each model component (i.e. sparsity, graph smoothness, and graph similarity) is discussed in this section. To analyze the effect of each component, each of the component weights (i.e. \( \alpha \), \( \beta \), and \( \gamma \)) changes independently while the other weights set to 1. Figs. 7, Fig. 8 and Fig. 9 represent the effectiveness of sparsity, graph smoothness, and graph similarity respectively. As can be seen from Fig. 7, changing the sparsity component weight results in lower change in the output result, which implies that in this scenario, sparsity is not a critical parameter. On the other hand, it seems that the changing graph similarity component weight has a considerable change in the output result (See Fig. 9), which means that this is the most effective component in this experiment.

The effect of eliminating each model component is also represented in Fig. 10. As it can be seen, omitting the graph similarity (i.e. \( \gamma = 0 \)) has the most devastating effect on the model performance, increased test number from 150 to 370, but still outperforms traditional CS because of the other two model components.
The node value sparsity has a minor effect and lacking this component makes only an increase of 20% in test number (i.e. an increase from 150 to 180).

**Conclusion**

In this paper, the contextual model-based CS is introduced to decrease the number of required tests for detection COVID-19. Despite the popularity of group testing, it was shown that this approach is not applicable in most real-world scenarios because of its strong dependency on data sparsity. Although the CS outperforms group testing, it also suffers from sparsity dependency which limits its efficiency in real scenarios. To our knowledge, no contextual information (i.e. age, underlying disease, symptoms) has been reported in the model-based CS.

In our CMB-CS, the context information is employed to build a model that helps recover CS samples more precisely with much fewer samples. The number of required tests for the accurate result (i.e. MSE < 0.01) is decreased more than four times. Moreover at the similar MSE, in the worst case, the required test is decreased at 68% compared to the traditional CS. In CMB-CS, three components are used to limit the search space: graph smoothness, node value sparsity and graph similarity. In other words, the recovered output is chosen from the search space which have the most smoothness, most sparse node and most similar to the reference graph. This optimization problem was formulated as a convex problem and solved by ADMM method.

The results show that among the available contextual information, age and fever have the most effect on the model performance. Although the “loss of consciousness” has a minor effect on our proposed model, it is shown that just using this information makes CMB-CS still superior to the traditional CS (i.e. the sampling rate decreased 33%).

Scrutinizing the effect of model components (i.e. smoothness, node value sparsity, and graph similarity) shows that graph similarity which employs contextual information has the most impact on the model performance (i.e. eliminating this component, increase the sampling rate about 36%). This indicates that our proposed model can be more useful in such cases that sparsity is not fully supported (i.e. the number of positive result is high). As a whole, results reveal the fact that integrating real-world contextual information such as underlying disease characteristics, symptoms, and social contact can improve the CS process in terms of both decreasing the number of tests and results accuracy.

Besides those mentioned advantages, our proposed model can be more beneficial in other aspects that can be extended in future works. Contrary to the group testing which can only produce binary results (i.e. positive or negative result), in CMB-CS the output can be any value. It means that in addition to infection status, the viral concentration can also be estimated if the appropriate dataset is available. Furthermore, the model can be customized to

**Table 1**

MSE, False positive and false negative results, traditional CS vs CMB-CS, at various test numbers.

| M (Number of tests) | Traditional CS | Contextual Model-based CS |
|---------------------|----------------|---------------------------|
|                     | MSE            | False Positive | False Negative | MSE            | False Positive | False Negative |
| 50                  | 0.2971         | 8              | 12             | 0.2469         | 8              | 9             |
| 60                  | 0.2956         | 10             | 9              | 0.2374         | 8              | 8             |
| 70                  | 0.2918         | 9              | 9              | 0.2235         | 9              | 6             |
| 80                  | 0.2893         | 8              | 9              | 0.2321         | 6              | 8             |
| 90                  | 0.2811         | 8              | 8              | 0.2055         | 7              | 6             |
| 100                 | 0.2780         | 6              | 9              | 0.1873         | 5              | 8             |
| 110                 | 0.2737         | 7              | 10             | 0.1558         | 4              | 8             |
| 120                 | 0.2714         | 6              | 10             | 0.1344         | 5              | 5             |
| 130                 | 0.2681         | 8              | 7              | 0.1151         | 6              | 3             |
| 140                 | 0.2528         | 8              | 7              | 0.0030         | 0              | 2             |
| 150                 | 0.2510         | 10             | 5              | 0.0003         | 1              | 0             |
| 160                 | 0.2478         | 9              | 5              | 0.0001         | 0              | 0             |
| 170                 | 0.2399         | 7              | 7              | 0.0000         | 0              | 0             |
| 180                 | 0.2306         | 6              | 8              | 0.0000         | 0              | 0             |
| 190                 | 0.2238         | 7              | 7              | 0.0000         | 0              | 0             |
| 200                 | 0.2112         | 9              | 5              | 0.0000         | 0              | 0             |

Fig. 6. The effect of each contextual information item in the output result separately.
Fig. 7. The effect of changing $\beta$, the sparsity component weight, on the model performance.

Fig. 8. The effect of changing $\alpha$, the graph smoothness component weight, on the model performance.
specific age groups, since the COVID-19 symptoms are different in children and adults [22] another future direction is to customize the CMb-CS for each age group. Since the epidemic behavior of COVID-19 changes during the time, the contextual information can be adapted based on the virus’s behavior at the time of data collection. Furthermore, the proposed model can also be extended to other contagious diseases. Finally, other contextual information such as geographical information and mobility can be embedded to make the model richer.

The measurement of contextual information is the key challenge of our work. Nearly all of this information, except for age, is binary values. In other words, the intensity and weakness of the parameter have not cared. On the other hand, some of this information is based on an individual expression which human error is inevitable. The reported data validation is another challenge. Recently, supported by Google and leading universities (e.g. Harvard, Oxford and John Hopkins), the data on individual COVID-19 cases are collected from different sources and made publicly available.
Having more details about this data (e.g., how long he/she in touch with a COVID-19 patient, fever degree, level of consciousness, etc.) can lead to a more precise model and makes more accurate results. Moreover, with the popularity of wearable health sensors, this information can be collected more accurately so the human error is eliminated.

The CMB-SC can be used in many public health monitoring scenarios. Additional to testing individuals, the proposed model can be applied in various health-related applications including prevalence estimation, plan to exit the lockdown, etc. In a broader view, our proposed model can be employed in any scenario that the number of measurements is the critical factor. MRI and CT imaging are other applications that can benefit from our method by reducing the required measurement and consequently, improves the patient experience and reduces X-ray harmful effects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1

Alternating Direction Method of Multipliers (ADMM) [19]

Consider the following convex optimization problem:

\[
\begin{align*}
\text{Minimize } & f(x) \\
\text{Subject to } & Ax = b
\end{align*}
\]

Where \( x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n} \) and \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex. The Lagrangian for solving this problem is:

\[
L(x, y) = f(x) + y^T(Ax - b)
\]

Generally, the optimization problem can be examined from two aspects: primal problem and dual problem. The solution for the dual problem provides a lower bound for the primal solution. The main advantage of this approach is that we can solve the problem regardless of how complex the primal problem is. The dual problem is easy to solve due to its concave minimization nature. Using dual function would also lead to an efficient and reliable solution. The dual problem is defined as:

\[
g(y) = \inf_{x} L(x, y) = -f^*(-A^Ty) - b^Ty
\]

Where \( f^* \) is a dual function of \( f \) [19]. Thus, the dual problem is: minimize \( g(y) \).

We can find a solution for the primal problem (\( x^* \)) from dual problem solution (\( y^* \)) as follows:

\[
x^* = \text{argmin}_x L(x, y^*)
\]

Several methods can be used to solve the above problem. The first one is the dual ascent that solves it using gradient ascent. The dual ascent method is implemented as follows:

\[
x^{k+1} = \text{argmin}_x L(x, y^k)
\]

\[
y^{k+1} = y^k + \alpha^k(Ax^{k+1} - b)
\]

Where \( \alpha^k > 0 \) the step size and \( k \) is the iteration counter. \( x \) and \( y \) are the primal and dual variable, respectively.

One of the main advantages of dual ascent is that it can be implemented in a decentralized form. In other words, \( f \) is separable in the form below:

\[
f(x) = \sum_{i=1}^N f_i(x_i)
\]

Where \( x = (x_1, x_2, \ldots, x_N) \) and \( x_i \) are subvector of \( x \). The matrix \( A \) also can be decomposed as follows:

\[
A = [A_1, \ldots, A_N]
\]

Hence, \( Ax = \sum_{i=1}^N A_i x_i \). The Lagrangian problem can be written as:

\[
L(x, y) = \sum_{i=1}^N L_i(x_i, y) = \sum_{i=1}^N \left( f_i(x_i) + y^T A_i x_i - \left( \frac{1}{N} \right) y^T b \right)
\]

In this way, the a-minimization step can be divided into \( N \) separate problems and solved in parallel as follows

\[
x^{k+1}_i = \text{argmin}_x L_i(x_i, y^k)
\]

\[
y^{k+1} = y^k + \alpha^k(Ax^{k+1} - b)
\]

Despite the merit of the dual ascent method, it has some limitations. The main disadvantage is that if \( f \) is a non-zero affine function of \( x \), then \( L \) is unbounded below in \( x \) and so the \( x \)-update fails.

The Augmented Lagrangian methods were introduced to add robustness to the dual ascent method and also reduce the limitation of \( f \). The Augmented Lagrangian is:

\[
L_p(x, y) = f(x) + y^T(Ax - b) + \left( \frac{\eta}{2} \right) \| Ax - b \|^2
\]

Where \( \eta > 0 \) is the penalty parameter. Applying The Augmented Lagrangian method yields the following algorithm:

\[
x^{k+1} = \text{argmin}_x L_p(x, y^k)
\]

\[
y = y + \rho (Ax^{k+1} - b)
\]

This form is known as the method of multipliers and it has been proven that it covers more general conditions. But this advantage comes at a cost. The \( L_p \) is not separable even if the \( f \) is. Hence, the \( x \) minimization step cannot be carried out in parallel to \( x \).

ADMM is the algorithm that benefits from both decomposability of dual ascent and good convergence of the method of multipliers. ADMM solves the problem of the form:

\[
\begin{align*}
\text{minimize } & f(x) + g(z) \\
\text{subject to } & Ax + Bz = c
\end{align*}
\]

Where \( x \in \mathbb{R}^n \) and \( z \in \mathbb{R}^m \) are both convex. As can be seen, in this form the variable has been split into two parts \( x \) and \( z \) and also the objective function is separable. The augmented Lagrangian is:

\[
L_p(x, z, y) = f(x) + g(z) + y^T (Ax + Bz - c) + \left( \frac{\eta}{2} \right) \| Ax + Bz - c \|_2^2
\]

The ADMM algorithm finds the optimal solution via iterations:

\[
x^{k+1} = \text{argmin}_x L_p(x, z^k, y^k)
\]

\[
z^{k+1} = \text{argmin}_z L_p(x^{k+1}, z, y^k)
\]

\[
y^{k+1} = y^k + \rho (Ax^{k+1} + Bz^{k+1} - c)
\]

In ADMM, \( x \) and \( z \) are updated sequentially. The minimization of each variable in separate steps is what allows decomposition when \( f \) or \( g \) is separable.

A1. ADMM scaled form

ADMM can be rewritten differently by combining the two latter terms in (38) and scaling the dual variable \( y \):

\[
y^T r + \left( \frac{\eta}{2} \right) \| r \|_2^2 = \frac{\eta}{2} \left( \| r + \frac{1}{\eta} y \|_2^2 - \frac{1}{\eta^2} \| y \|_2^2 \right) = \frac{\eta}{2} \| r + u \|_2 - \left( \frac{\eta}{2} \right) \| u \|_2^2
\]
Where $u = \frac{1}{\eta} y$ is a scaled dual variable and $r = Ax - b$. We can express the scaled form of ADMM as follows:

$$x^{k+1} := \text{argmin}_x \left( f(x) + \frac{\eta}{2} \|Ax + Bz^k - c + u^k\|^2_2 \right)$$

(44)

$$z^{k+1} := \text{argmin}_z \left( g(z) + \frac{\eta}{2} \|Ax^{k+1} + Bz - c + u^k\|^2_2 \right)$$

(45)

$$u^{k+1} := u^k + Ax^{k+1} + Bz^{k+1} - c$$

(46)

Where $u^k$ is the cumulative residual:

$$u^k = u^0 + \sum_{j=1}^{k} Ax^j + Bz^j - c$$

(47)

References

[1] D.B. Vinh, et al., Overview of COVID-19 testing and implications for otolaryngologists, Head Neck 49 (7) (2020) 1629–1633.

[2] A. Biosystems, Real-time PCR handbook, Life Technologes Carlsbad, 2015 CA, USA.

[3] K. He, et al., Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images, Pattern Recognit. 113 (2021) 107828.

[4] X. Chen, L. Yao, T. Zhou, J. Dong, Y. Zhang, Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images, Pattern Recognit. 113 (2021) 107828.

[5] R. Dorfman, The detection of defective members of large populations, Ann. Math. Stat. 14 (4) (1943) 436–440.

[6] A. Shlomai, H. Ben-Zvi, A.G. Bendersky, N. Shafran, E. Goldberg, E.H. Sklan, Nasopharyngeal viral load predicts hypoxemia and disease outcome in admitted COVID-19 patients, Crit. Care 24 (1) (2020) 1–3.

[7] Y. Wang, et al., Kinetics of viral load and antibody response in relation to COVID-19 severity, J. Clin. Invest. 130 (10) (2020) 5235–5244.

[8] R. Baraniuk, M.A. Davenport, M.F. Duarte, C. Hegde, An introduction to compressive sensing, Connexions (2011) [Online]. Available: http://cnx.org/content/col11313/1.57/.

[9] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser Basel, New York, NY, 2013.

[10] G.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Inf. Theory 56 (4) (2010) 1982–2001.

[11] J.N. Eberhardt, N.P. Breuckmann, C.S. Eberhardt, Multi-stage group testing improves efficacy of large-scale COVID-19 screening, J. Clin. Virol. 128 (2020) 104382.

[12] A. Sandryhaila, J.M.F. Moura, Discrete signal processing on graphs: frequency analysis, IEEE Trans. Signal Process. 62 (12) (2014) 3042–3054.

[13] M. Daković, L. Stanković, E. Sejdić, Local smoothness of graph signals, Math. Probl. Eng. 2019 (2019) 1–14.

[14] W. Lin, J. Huang, C.Y. Suen, L. Yang, A feature extraction model based on discriminative graph signals, Expert Syst. Appl. 139 (2020) 112861.

[15] F. Dornaika, L. Wang, Sparse graphs with smoothness constraints: application to dimensionality reduction and semi-supervised classification, Pattern Recognit. 95 (2019) 285–295.

[16] S. Chen, A. Sandryhaila, J.M.F. Moura, J. Kováč, Signal recovery on graphs: variation minimization, IEEE Trans. Signal Process. 63 (17) (2015) 4609–4624.

[17] X.-L. Hu, J. Wen, Z. Lai, W.X. Wong, L. Shen, Binary sparse signal recovery algorithms based on logic observation, Pattern Recognit. 90 (2019) 147–160.

[18] D. Ge, X. Jiang, Y. Ye, A note on the complexity of L p minimization, Math. Program. 129 (2) (2011) 285–299.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers, vol. 3, no. 1. Now Publishers, Inc., 2011.

[20] Y. Liu, T. Liu, J. Liu, C. Zhu, Smooth robust tensor principal component analysis for compressed sensing of dynamic MRI, Pattern Recognit. 102 (2020) 107252.

[21] KarimyMahmoodAabanMarziehMesiMehdiArmoobBahramKoohestani, HamidRezaAzani, Data on epidemiological and clinical characteristics of coronavirus-infected (COVID-19) individuals in Iran. figshare. Dataset, Hadi (2020) https://doi.org/10.6084/m9.figshare.12446120.v1.

[22] N.C. Davies, et al., Age-dependent effects in the transmission and control of COVID-19 epidemics, Nat. Med. 26 (8) (2020) 1205–1211.

[23] Global Health, a data science initiative," 2021. [Online]. Available: https://globalhealth/, accessed on 20 May, 2021

Mohammad Khansari received his B.S, M.S and PhD degrees in Computer Engineering all from Sharif University of Technology, Tehran, Iran, in 1996, 1998 and 2008 respectively. He had a short-time research fellowship from DAAD, Germany at IPSI-Fraunhofer research institute. He served as the faculty member of Sharif International campus, Kish Island, director of Iran Telecommunication Research Center (ITRC) and director of IT and Cyberspace Center of University of Tehran. Currently, he is the faculty member of Faculty of New Sciences and Technologies (Associate Professor), University of Tehran. His main research interests are network science and complex networks, multimedia IoT, and wireless multimedia sensor networks.

Mehdi Hasanasab received his B.S, M.S in electrical and mechatronic engineering respectively from Qazvin Azad University (QAU). He is now PhD student in University of Tehran. His main research field is data compression methods, signal reconstruction and graph signal processing.